

ADVANCES IN PHYSICS

A QUARTERLY SUPPLEMENT
of the
PHILOSOPHICAL MAGAZINE

EDITOR

PROFESSOR N. F. MOTT, M.A., D.Sc., F.R.S.

EDITORIAL BOARD

SIR GEORGE THOMSON, M.A., D.Sc., F.R.S.

PROFESSOR A. M. TYNDALL, C.B.E., D.Sc., F.R.S.

SIR LAWRENCE BRAGG, O.B.E., M.C., M.A., D.Sc., F.R.S.

VOLUME 1

JANUARY 1952

NUMBER 1

PRICE per part 15s. 0d.

PRICE per annum £2 15s. 0d. post free

PRINTED AND PUBLISHED BY TAYLOR & FRANCIS LTD.

RED LION COURT, FLEET ST., LONDON E.C.4

Atomic Scientists' News

Journal of the Atomic Scientists' Association

President

Executive Vice-President

Vice-Presidents:

Sir WALLACE AKERS.

Professor LORD CHERWELL, F.R.S.

Professor H. S. W. MASSEY, F.R.S.

Professor N. F. MOTT, F.R.S.

Professor R. E. PEIERLS, F.R.S.

Professor F. A. PANETH, F.R.S.

Professor H. W. B. SKINNER, F.R.S.

Professor P. M. S. BLACKETT, F.R.S.

General Sec.: Dr. J. L. MICHIELS.

Professor M. H. L. PRYCE, F.R.S.

Professor KATHLEEN LONSDALE, F.R.S.

Sir JOHN COCKCROFT, F.R.S.

Sir CHARLES DARWIN, F.R.S.

Professor P. B. MOON, F.R.S.

Professor M. L. E. OLIPHANT, F.R.S.

Professor F. E. SIMON, F.R.S.

Professor Sir GEOFFREY TAYLOR, F.R.S.

Professor Sir GEORGE THOMSON, F.R.S.

Professor C. F. POWELL, F.R.S.

Treasurer: Dr. L. E. J. ROBERTS.

- The Atomic Scientists' Association is an association of scientists whose work has given them special knowledge of the consequences for the world of the use and misuse of atomic energy. To make known the true facts about atomic energy and its implications, it publishes every two months the **Atomic Scientists' News**.
- Full membership of the Association is open to all scientists able to put before the public an informed opinion upon some aspect of atomic energy. Others interested in its work may become associate members. Enquiries c/o Professor J. Rotblat, Physics Department, St. Bartholomew's Hospital, London, E.C.1.
- Commencing September 1951, the **News** will be published on behalf of the Association by Messrs. Taylor & Francis, Ltd.

Vol. 1

NOVEMBER 1951

No. 2

CONTENTS

Editorial

The British Experimental Pile - - - - - M. o. S.

The Isotope Conference at Harwell - - - - - Prof. J. Rotblat

Pure Science - - - - - Sir John Cockcroft, C.B.E., F.R.S.

A Quaker Scientist Visits the Moscow

Academy of Sciences - - - - - Prof. K. Lonsdale, F.R.S.

Press Review - - - - - Dr. F. R. N. Nabarro

Letter to the Editor - - - - - Prof. H. W. B. Skinner, F.R.S.

Book Reviews - - - - - By Dr. D. F. Bracher, Dr. G. Stephenson
and Dr. G. O. Jones

Price to non-members 6/-

(plus postage)

Annual subscription £1 12s. 6d.

(post free)

Non-members apply to the Printers and Publishers:—

Messrs. TAYLOR & FRANCIS, LTD.

Red Lion Court, Fleet Street, London, E.C.4

CONTENTS

The Mean Free Path of Electrons in Metals. By E. H. SONDHEIMER, Royal Society Mond Laboratory, Cambridge	1
On the Generation of Vacancies by Moving Dislocations. By F. SEITZ, University of Illinois, Urbana, Ill., U.S.A.	43
Crystal Growth and Dislocations. By F. C. FRANK, H. H. Wills Physical Laboratory, University of Bristol	91

5856-22

EDITORIAL

A NEW venture in scientific publishing, even though it appears as a supplement to a journal so long established as the *Philosophical Magazine*, requires an editorial to explain its purpose. The purpose of 'Advances in Physics' is to publish articles, written by experts for experts, which will review whatever is of current interest in physics. The range of subjects on which articles will be acceptable will be as wide as for the *Philosophical Magazine*; that is to say it will include theoretical and experimental physics and certain branches of applied mechanics; but, in the interests of the reader whose interests are more limited, it is hoped that whenever possible the contributions in each number will be confined to a more restricted field. This number, for example, deals with some problems in the physics of the solid state, while further issues are planned which will deal with the properties of liquid helium, the theory of dislocations and the science of the upper atmosphere. Research workers who do not subscribe to the supplement may, therefore, find it worth while to buy occasional individual numbers. But, though the subject matter of each number is to be drawn from a rather narrow circle, the editors hope that the contributors will be drawn from a circle as wide as possible; they hope that the articles in 'Advances in Physics' will represent the international nature of science, and are therefore particularly glad to welcome among the papers already accepted several contributors from other countries.

'Advances in Physics' will appear quarterly, and this perhaps will give it its character and distinguish it from other journals in this country which appear annually. The aim of the journal is to publish specialist articles, of admittedly ephemeral value, in rather narrow fields, and to publish them quickly. When some advance is made, in which one or more laboratories are involved, and which has been described to the world

by letters and by papers in scattered journals, the editors hope to induce an expert in the field to put on paper a properly connected and critical account of what he and his colleagues have done. An article in 'Advances in Physics' may serve in some cases as the stepping-stone, the half-way house, between the series of papers in the *Philosophical Magazine* or elsewhere in which the results of research are first recorded and the monograph or book in which they are given their final and well-considered description.

The articles in the present number may be cited as examples of the purpose which the editors have in mind. They are all welcome ; but, in the interests of brevity, one must be chosen. Let us take that by Dr. Frank on Crystal Growth and Dislocations. Here is a subject on which the author of this article has produced a theory which looks as if it might be the answer to many puzzling questions. Experiments carried out at the Royal Holloway College, London, at Glasgow University and also at Bristol have been correlated with the theory, and as a result, within the last few months, many of the main features of the subject have become clear. The subject was therefore ready for a review.

The editors would at all times welcome suggestions and offers to contribute articles. If they know in advance that an article is coming, they can provide quick publication. If articles are submitted in languages other than English, they will, if possible, arrange publication. The editors hope that the supplement will fill a need in the scientific world, and in doing so contribute to the Advance of Physics.

ADVANCES IN PHYSICS

A QUARTERLY SUPPLEMENT

of the

PHILOSOPHICAL MAGAZINE

VOLUME 1

JANUARY 1952

NUMBER 1

The Mean Free Path of Electrons in Metals

By E. H. SONDHEIMER*,

Royal Society Mond Laboratory, Cambridge†

§1. INTRODUCTION

1.1. The foundations of the modern electron theory of metals were laid at the beginning of the present century, when the existence of a gas of free electrons was postulated by Drude in order to explain the conducting properties of metals; the behaviour of the electrons was subsequently analysed by Lorentz by means of the statistical methods of the dynamical theory of gases. The chief success of the Drude-Lorentz theory was the prediction of the Wiedemann-Franz law connecting the electrical and thermal conductivities, but later developments revealed an increasing number of serious difficulties, outstanding among them the inability of the theory to explain why the conduction electrons do not contribute appreciably to the specific heat of a metal. This paradox was not resolved until the advent of quantum mechanics, when Pauli and Sommerfeld applied the Fermi-Dirac statistics to the free electrons in a metal, showing that in this way most of the contradictions could be reconciled.

The Drude-Lorentz-Sommerfeld theories are essentially formal in character. They involve as arbitrary parameters the number n of free electrons per unit volume, which is assumed to be of the same order as the number of atoms per unit volume, and the mean free path l of the electrons which is to be determined from a comparison of theory and experiment. So far as the electrical conductivity σ_0 is concerned, the results of the Sommerfeld theory are summarized by the formulae

$$n = \frac{8\pi}{3} \left(\frac{m\bar{v}}{h} \right)^3, \quad \dots \quad (1a), \quad \sigma_0 = \frac{n\epsilon^2 l}{m\bar{v}}, \quad \dots \quad (1b)$$

in which h is Planck's constant, $-\epsilon$ is the charge and m is the mass of an electron, and \bar{v} is the velocity of an electron at the surface of the Fermi distribution. σ_0 therefore depends on n , l and fundamental constants only. In order to obtain the correct order of magnitude for

* Fellow of Trinity College, Cambridge.

† Now at the Department of Mathematics, Imperial College, London.

the conductivity and the correct temperature variation, it is necessary to assume that l is of the order of several hundred interatomic distances at ordinary temperatures, and increases rapidly in a pure metal towards very low temperatures.

Such long free paths which vary with temperature are very difficult to explain on classical theory, but they can be understood on the basis of the recent developments in the theory of metals. These were initiated by Bloch and elaborated by a great many authors; they have been concerned with a detailed quantum-mechanical analysis of the motion of electrons in a crystal lattice, and they have made it possible to give precise meanings to the two fundamental concepts of the 'number of free electrons' and the 'mean free path' and to obtain numerical estimates of these quantities in certain cases. The electrons in a metal are regarded as distributed over a number of energy bands, filling most of them completely. All the electrons are free to move through the lattice, but only those which are contained in incompletely filled energy bands can contribute to the resultant current and are to be regarded as free electrons for the purposes of conduction theory. The number of free electrons in a metal is of the same order as the number of atoms: the precise number, however, depends on the detailed configuration of the energy bands and need not be a simple multiple or submultiple of the number of atoms. The variation of n with temperature is negligible, since at ordinary temperatures the free electrons form a highly degenerate Fermi-Dirac gas.

An electron can move freely through a perfect and rigid crystal lattice and there is no resistance. In a pure metal a finite free path is caused by the thermal vibrations of the lattice and is of the same order as the average wavelength of the sound waves in the metal, which is large compared with the interatomic distance and is increased by lowering the temperature. The free path does not increase indefinitely, however, as the temperature is lowered, and at very low temperatures it tends to a constant 'residual' value l_r , which is determined by static lattice imperfections such as the presence of impurity atoms, and which is of the order of the distance between the impurities.

The following theoretical formula may be deduced for the electrical conductivity of a metal, subject to many simplifying assumptions concerning the interaction between the electrons and the lattice vibrations (see, for example, Wilson 1936, Chapter VI):

$$\frac{1}{\sigma_0} = \frac{m\bar{v}}{n\epsilon^2 l_r} + \left(\frac{m}{2}\right)^{1/2} \frac{9\pi\hbar^2 C^2}{8n\Delta\epsilon^2 M k \Theta \zeta^{3/2}} \left(\frac{T}{\Theta}\right)^5 \int_0^{\Theta/T} \frac{z^5 dz}{(e^z - 1)(1 - e^{-z})}. \quad (2)$$

Here k is Boltzmann's constant, ζ is the Fermi energy level ($\zeta = \frac{1}{2} m \bar{v}^2$), Θ is the Debye temperature, M is the mass of an atom, Δ is the volume of the unit cell, and C is a constant which determines the interaction between the electrons and the lattice. According to this formula the 'ideal' and 'residual' resistances are additive, and the ideal resistance is proportional to T at high and to T^5 at very low temperatures. (For

commercially pure metals the residual resistance can be neglected at ordinary temperatures.) The temperature variation of the resistance predicted by (2) is, generally speaking, in good agreement with observation, particularly for the monovalent metals. The absolute value of the free path at any temperature can be estimated by combining (1) and (2) and substituting reasonable values for the parameters. The chief uncertainty concerns the magnitude of the interaction constant C , estimates of which can be obtained by numerical integration if the wave functions of the conduction electrons are known; it is of the same order as the Fermi energy ζ . The estimates of the free path are of the right order of magnitude, but precise numerical values cannot be obtained in this way.

1.2. An important conclusion to be drawn from the detailed quantum-mechanical theory is that the simple Sommerfeld treatment of the conduction phenomena remains correct within certain limits. In the Sommerfeld theory the electrons are regarded as perfectly free, their energy being proportional to the square of the velocity; this remains approximately true for electrons moving in a lattice, but the mass m which appears in equations (1) and (2) must be regarded as an effective mass which is of the same order of magnitude as, but not necessarily equal to, the mass of a free electron. This model of 'quasi-free' electrons, which is implicit in the derivation of (2), applies most closely to the monovalent metals in which the conduction electrons are all contained in a single energy band; for these metals, moreover, the number of conduction electrons should be precisely one per atom. In multivalent metals, in which the electrons occupy more than one band, the model may still be used to give a semi-quantitative description of the simpler conduction phenomena, but quantities such as n and m must then be regarded as representing certain averages of the numbers of electrons and the effective masses of the electrons in the various bands, and the precise numerical values have no immediate physical significance. It would, of course, be possible to consider more complicated models, for example one in which the conduction electrons are contained in two overlapping energy bands. In the present state of the theory of metals, however, it is impossible to work out a theory which fully takes into account the electronic structure peculiar to any particular metal, and instead of introducing a large number of parameters of doubtful physical significance it is best to work with the simplest model which gives reasonable results. For more complicated conduction phenomena, however, particularly those which are associated with the presence of a magnetic field or with anisotropy effects, the free-electron model is entirely inadequate (it leads, for example, to a zero magneto-resistance effect), and even a qualitative theory can only be obtained by using a model in which the energy surfaces do not form a singly-connected set of spheres.

In the Sommerfeld theory the free path is most conveniently introduced through the time of relaxation τ , which is defined as follows. Let $\mathbf{v} = (v_x, v_y, v_z)$ be the velocity of an electron, and let $2(m/h)^3 f(\mathbf{v}, \mathbf{r}) d\mathbf{r} d\mathbf{v}$

however, possible to obtain l directly by measuring the conductivity under conditions where the free path may be compared with some other characteristic length in the metal. Effects of this type have attracted much attention in the last few years, and it is with them that we shall henceforth be concerned.

The most obvious method is to use a thin film or wire and to arrange that the free path is comparable in magnitude with the thickness or diameter of the specimen; the artificial limitation of the free path by the boundaries of the specimen causes an increase in the resistivity above its value in the bulk metal, and this may be used to deduce the ratio of free path to thickness or diameter. This topic is reviewed in §2.

In §3 we consider the more complicated effects which occur when a thin specimen is placed in a magnetic field. The ordinary bulk magneto-resistance effect depends in a complicated way on the binding of the electrons in the lattice and is zero for quasi-free electrons; in thin specimens, however, where boundary scattering of electrons is important, the alteration of the electron trajectories in a magnetic field in general leads to a non-zero change of resistance even if the electrons are regarded as free. These curious 'geometrical' effects, being classical in nature, are entirely different from the bulk effect and are essentially simpler to understand. The details of the phenomena vary with the shape of the specimen and with the relative configurations of specimen, current and magnetic field; their analysis in all cases involves a new quantity with the dimensions of length, namely the radius $r_0 = m\bar{v}c/\epsilon H$ of a free-electron orbit in a magnetic field H . The experiments therefore give, in addition to the free path, a direct estimate of the momentum $m\bar{v}$ of the electrons at the surface of the Fermi distribution, and hence of the number of free electrons according to equation (1*a*). The method is, however, severely restricted in practice by the disturbing effect of the bulk magneto-resistance phenomenon.

In §4, finally, we consider the so-called anomalous skin effect in metals. This is a more sophisticated size effect in which the free path is compared, not with the physical dimensions of the specimen, but with the distance to which a high-frequency electric field penetrates into the metal. Experiments on the high-frequency skin resistance of metals allow values of l/δ to be deduced, where δ is the classical skin penetration depth, and since δ depends only on the frequency and the d.c. electrical conductivity, the free path can again be obtained directly from experimental magnitudes.

1.4. These phenomena present interesting problems from both experimental and theoretical points of view. Let us consider the orders of magnitude of the various characteristic lengths. The free path in metals at room temperatures is of the order of 10^{-5} cm or less, but in a pure metal at liquid-helium temperatures the high values of the conductivity indicate that it may be as large as 10^{-2} cm. It is clear, therefore, that extremely thin films would be required in experiments on the size effect at normal temperatures; such films are difficult to prepare and usually show

subsidiary resistive effects, often time dependent, which tend to obscure the pure geometrical limitation of the free path with which we are alone concerned. In order to obtain results which are free from ambiguity, it is therefore essential to carry out the experiments at very low temperatures where relatively large specimens (of thickness $\sim 10^{-3}$ cm) may be used. The necessity for using specimens of this order of thickness becomes even greater when we consider the effects which take place in a magnetic field. These effects show up when the orbit radius r_0 is comparable with the thickness of the specimen. Since \bar{v} is of the order of 10^8 cm/sec, a magnetic field of reasonable magnitude (several kilogauss) corresponds to $r_0 \sim 10^{-3}$ cm; and since r_0 is inversely proportional to H , impracticably large magnetic fields would be required in the case of films much thinner than this. Finally, in the anomalous skin effect l must be large compared with δ : the ratio l/δ is proportional to ν^2 and to the square root of the frequency, so that both very low temperatures and high frequencies are required. For a pure metal at liquid-helium temperatures and for microwave frequencies, l/δ is of the order of 100.

One may say, therefore, that the size effects are essentially low-temperature phenomena. Since the free path at low temperatures varies from specimen to specimen and is not a characteristic property of the metal, it is usual to measure (or estimate in some way) the value of the bulk conductivity σ_0 which corresponds to the free path l , and to express the results of the experiments in terms of the ratio $\sigma_0 l$. This ratio determines directly the number of conduction electrons per unit volume, since, according to equations (1 a) and (1 b),

$$\frac{\sigma_0}{l} = \frac{n e^2}{m \bar{v}} = \left(\frac{8\pi}{3} \right)^{1/3} \frac{e^2 n^{2/3}}{h} = 7.1 \times 10^7 n^{2/3}. \quad (4)^*$$

Note that this relation involves only the assumption that the electrons are quasi-free, but not that their effective mass is necessarily equal to the mass of a free electron.

1.5. Many approximate theoretical treatments have been given of the various free-path phenomena, but we shall not refer to these (except where they are particularly relevant to the discussion), as they have been superseded by the strict statistical analysis based on the Boltzmann equation for the distribution function of the conduction electrons†. This equation is formed by equating the rate of change in f due to external fields to the rate of change due to the collision mechanism, which is assumed to be given by equation (3). In the presence of an electric field \mathbf{E} and a magnetic field \mathbf{H} , the Boltzmann equation for quasi-free electrons takes the form

$$-\frac{e}{m} \left(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H} \right) \cdot \text{grad}_{\mathbf{v}} f + \mathbf{v} \cdot \text{grad}_{\mathbf{r}} f = -\frac{f - f_0}{\tau}, \quad (5)$$

* Gaussian units are used throughout the present article.

† The use of the Boltzmann equation in this connection appears to have been first suggested by Peierls.

which is purely classical except that the mass m is to be regarded as an effective mass, while the equilibrium distribution function is the Fermi-Dirac function

$$f_0(E) = \frac{1}{e^{(E-\zeta)/kT} + 1} \quad \dots \quad (6)^*$$

The time of relaxation τ is supposed to depend on the absolute value of \mathbf{v} only.

f is a function of the velocity vector \mathbf{v} and the space vector \mathbf{r} , and the analysis of the size effects depends essentially on the appearance in equation (5) of the term involving the space derivatives of f , which is required in the bulk metal only if temperature gradients are present. This term takes into account the non-uniform distribution in space of the conduction electrons which is characteristic of the size-effect phenomena. A general feature of the analysis is that boundary conditions have to be imposed on the solutions of the Boltzmann equation; these are determined in each case by the nature of the scattering at the physical boundaries of the specimen.

In treating the various problems in this way we may apply the results directly to the alkali metals, in which the electrons are most nearly free, and with somewhat lesser confidence to the noble metals; for more complex metals the model will provide only a qualitative picture, but should give the orders of magnitude correctly. In the presence of a magnetic field, corrections of uncertain amount must be applied to the classical formulae to allow both for the bulk magneto-resistance effect and, possibly, for the effect of the quantization of the electron orbits in a magnetic field. These refinements are difficult to take into account in any satisfactory way; the most reasonable procedure, therefore, is to neglect them entirely and to restrict the experiments to metals in which they are of least importance (see § 3.1).

The main problem, then, is to solve equation (5) for the various cases of interest, and to use the solution to calculate the current density \mathbf{J} by means of the usual formula

$$\mathbf{J} = -2e \left(\frac{m}{h} \right)^3 \int \mathbf{v} f d\mathbf{v} \quad \dots \quad (7)$$

1.6. Objections have been raised from time to time against the whole basis of the theory of metallic conductivity: thus the uncertainty principle has been used to argue that the uncertainty in the energy of a single electron, which is determined by the magnitude of the time of relaxation, may be so large in some cases as to render meaningless any description of the situation by means of a distribution function which obeys the classical Boltzmann equation (for a detailed discussion of this somewhat oversimplified statement compare, for example, Peierls 1934). These objections have never been satisfactorily met, and it must be

* E here denotes the energy of an electron, but elsewhere in this article E stands for electric field strength.

2.111. We can now proceed with the calculation of the current density $J(z)$. Combining (7), (8), (11), introducing polar coordinates (v, θ, ϕ) in the \mathbf{v} -space (with $v_z = v \cos \theta$), and remembering that f_0 depends on $v = |\mathbf{v}|$ only, we obtain

$$J(z) = -\frac{2e^2 m^2 E}{h^3} \int_0^\infty dv \int_0^{2\pi} d\phi \tau v^3 \cos^2 \phi \frac{\partial f_0}{\partial v} \\ \times \left[\int_0^{\frac{1}{2}\pi} \sin^3 \theta \left\{ 1 - \exp \left(-\frac{z}{\tau v \cos \theta} \right) \right\} d\theta \right. \\ \left. + \int_{\frac{1}{2}\pi}^\pi \sin^3 \theta \left\{ 1 - \exp \left(\frac{a-z}{\tau v \cos \theta} \right) \right\} d\theta \right]. \quad (12)$$

The integration over ϕ is trivial, and the integration over v is carried out by means of the formula

$$-\int_0^\infty \psi(v) \frac{\partial f_0}{\partial v} dv = \psi(\bar{v}), \quad (13)$$

which holds for a degenerate electron gas. This gives, after some rearrangement,

$$J(z) = \frac{4\pi e^2 m^2 \tau \bar{v}^3}{h^3} E \int_0^{\frac{1}{2}\pi} \sin^3 \theta \left\{ 1 - \exp \left(-\frac{a}{2l \cos \theta} \right) \cosh \left(\frac{a-2z}{2l \cos \theta} \right) \right\} d\theta, \quad (14)$$

where $l = \tau \bar{v}$ is the free path of the electrons at the surface of the Fermi distribution.

Equation (14) gives the current distribution across the thickness of the film. For comparison with experiment, however, we require the apparent *overall* conductivity of the film, and we must therefore average the current density over all values of z from 0 to a . Carrying out the integration over z , we obtain for the effective conductivity

$$\sigma = \frac{1}{Ea} \int_0^a J(z) dz = \sigma_0 \left[1 - \frac{3l}{2a} \int_0^{\frac{1}{2}\pi} \sin^3 \theta \cos \theta \left\{ 1 - \exp \left(-\frac{a}{l \cos \theta} \right) \right\} d\theta \right], \quad (15)$$

σ_0 being the conductivity of the bulk metal, which is given by equation (1 b). The ratio of the resistivity $1/\sigma$ of the film to that of the bulk metal $1/\sigma_0$ is most conveniently written in the form, obtained from (15) by means of a simple substitution,

$$\frac{\sigma_0}{\sigma} = \frac{\Phi(\kappa)}{\kappa}, \quad (16)$$

where $\kappa = a/l$ and where

$$\frac{1}{\Phi(\kappa)} = \frac{1}{\kappa} - \frac{3}{8\kappa^2} + \frac{3}{2\kappa^2} \int_1^\infty \left(\frac{1}{t^3} - \frac{1}{t^5} \right) e^{-\kappa t} dt. \quad (17)$$

Repeated integration by parts gives the alternative expression

$$\frac{1}{\Phi(\kappa)} = \frac{1}{\kappa} - \frac{3}{4} \left(1 - \frac{1}{12} \kappa^2 \right) Ei(-\kappa) - \frac{3}{8\kappa^2} (1 - e^{-\kappa}) - \left(\frac{5}{8\kappa} + \frac{1}{16} - \frac{\kappa}{16} \right) e^{-\kappa}. \quad (18)$$

where

$$-Ei(-u) = \int_u^\infty \frac{e^{-t}}{t} dt, \quad (19)$$

which is a tabulated integral; this form of the result is particularly convenient for computation.

The limiting form for large κ (thick films) is

$$\frac{\sigma_0}{\sigma} = 1 + \frac{3}{8\kappa} \quad (\kappa \gg 1), \quad (20)$$

and for very thin films we have

$$\frac{\sigma_0}{\sigma} = \frac{4}{3\kappa \log(1/\kappa)} \quad (\kappa \ll 1). \quad (21)$$

These formulae are discussed further in § 2.3.

2.12. A somewhat more general theory, which does not assume that the scattering at the surface of the film is entirely diffuse, can be obtained as follows. We assume that a fraction p of the electrons is scattered elastically at the surface with reversal of the velocity component v_z , while the rest are scattered diffusely with complete loss of their drift velocity. p is supposed to be a constant independent of the direction of motion of the electrons. This is of course a highly artificial model, which in effect merely interpolates between the extreme cases of perfectly diffuse reflection, considered above, and perfectly specular reflection, for which the conductivity is unaltered. However, in the absence of any detailed theory of the nature of the surface scattering mechanism it is best to work with the simplest possible assumptions.

The distribution function of the electrons leaving the surface $z=0$ is now given by

$$f_0 + f_1^+(v_z, z=0) = p\{f_0 + f_1^-(-v_z, z=0)\} + (1-p)f_0, \quad (22)$$

and similarly, at $z=a$,

$$f_0 + f_1^-(v_z, z=a) = p\{f_0 + f_1^+(-v_z, z=a)\} + (1-p)f_0. \quad (23)$$

These equations are sufficient to determine $F(\mathbf{v})$, and instead of (11) we obtain for the distribution function

$$\left. \begin{aligned} f_1^+(\mathbf{v}, z) &= \frac{\epsilon \tau E}{m} \frac{\partial f_0}{\partial v_x} \left\{ 1 - \frac{1-p}{1-p \exp(-a/\tau v_z)} \exp\left(-\frac{z}{\tau v_z}\right) \right\} & (v_z > 0), \\ f_1^-(\mathbf{v}, z) &= \frac{\epsilon \tau E}{m} \frac{\partial f_0}{\partial v_x} \left\{ 1 - \frac{1-p}{1-p \exp(a/\tau v_z)} \exp\left(\frac{a-z}{\tau v_z}\right) \right\} & (v_z < 0). \end{aligned} \right\} \quad (24)$$

The current density is calculated as before, and the result is that the function $\Phi(\kappa)$ defined by (17) must be replaced by $\Phi_p(\kappa)$, where

$$\frac{1}{\Phi_p(\kappa)} = \frac{1}{\kappa} - \frac{3}{2\kappa^2}(1-p) \int_1^\infty \left(\frac{1}{t^3} - \frac{1}{t^5} \right) \frac{1-e^{-\kappa t}}{1-pe^{-\kappa t}} dt. \quad (25)$$

This reduces to (17) when $p=0$, and to the bulk-metal value $1/\kappa$ when $p=1$. It is, of course, again possible to write down alternative forms of (25) more convenient for numerical work; compare, for example, equation (31). Instead of (20) and (21), we have

$$\frac{\sigma_0}{\sigma} = 1 + \frac{3}{8\kappa}(1-p) \quad (\kappa \gg 1) \quad (26)$$

$$\text{and} \quad \frac{\sigma_0}{\sigma} = \frac{4}{3} \frac{1-p}{1+p} \frac{1}{\kappa \log(1/\kappa)} \quad (\kappa \ll 1). \quad (27)^*$$

Values of σ_0/σ for $p=0$ and $\frac{1}{2}$ are shown in Table 1.

TABLE 1

The resistivity of thin metallic films divided by the resistivity of the bulk metal

κ (thickness/free path)	σ_0/σ	
	$p=0$	$p=\frac{1}{2}$
0.001	182	73.5
0.002	100.4	41.5
0.005	46.6	20.0
0.01	26.5	11.8
0.02	15.3	7.1
0.05	7.69	3.87
0.1	4.72	2.62
0.2	3.00	1.91
0.5	1.90	1.402
1	1.462	1.206
2	1.221	1.102
5	1.081	1.039
10	1.0390	1.0191
20	1.0191	1.0095
50	1.0076	1.0038
100	1.0038	1.0019

2.2. *Thin wires.* The corresponding analysis for the case of wires has been carried out by MacDonald and Sarginson (1950), who considered a wire of square cross-section, and by Dingle (1950) for the more important but more difficult case of a wire of circular cross-section. The method is essentially the same as that for a film, but the details of the analysis, for which the reader is referred to the original papers, are more complicated because the problems are now two-dimensional and the Boltzmann

* The corresponding formula given by Fuchs (equation (23) of his paper) is incorrect.

equation takes the form of a *partial* differential equation in the space variables. The result for a cylindrical wire of diameter a is*, for $p=0$,

$$\frac{\sigma_0}{\sigma} = \frac{\Psi(\kappa)}{\kappa}, \quad (28)$$

where $\kappa=a/l$,

$$\frac{1}{\Psi(\kappa)} = \frac{1}{\kappa} - \frac{12}{\pi\kappa} \int_0^1 (1-t^2)^{1/2} S_4(\kappa t) dt, \quad (29)$$

and

$$S_n(u) = \int_1^\infty e^{-ut}(t^2-1)^{1/2} t^{-n} dt. \quad (30)$$

The result for $p \neq 0$ may be derived from that for $p=0$ by means of the simple relation

$$\left(\frac{\sigma}{\sigma_0}\right)_{\kappa, p} = (1-p)^2 \sum_{n=1}^\infty n p^{n-1} \left(\frac{\sigma}{\sigma_0}\right)_{n\kappa, p=0}; \quad (31)$$

this holds incidentally also for the case of a thin film, as may be shown by expanding (25) in ascending powers of p and comparing with (17).

For very thick and very thin wires the results reduce to

$$\frac{\sigma_0}{\sigma} = 1 + \frac{3}{4\kappa} (1-p) \quad (\kappa \gg 1) \quad (32)$$

and

$$\frac{\sigma_0}{\sigma} = \frac{1-p}{1+p} \frac{1}{\kappa} \quad (\kappa \ll 1). \quad (33)$$

(32) holds also for a square wire (κ being a/l as before where a is now the side of the wire), but (33) must be replaced by

$$\frac{\sigma_0}{\sigma} = \frac{1-p}{1+p} \frac{0.897}{\kappa} \quad (\kappa \ll 1). \quad (34)$$

Numerical values of σ_0/σ for a cylindrical wire are shown in Table 2.

2.3. *Discussion of the formulae.* It will be noted that, although the general theory is complicated, relatively simple formulae are obtained for specimens which are either thick or thin compared with the free path. The results for these limiting cases in fact possess a simple interpretation.

2.31. The results given above for *thick* specimens are particular cases of a formula applicable to wires with cross-sections of arbitrary shape, the general form of which follows from simple dimensional considerations (Dingle 1950). When the thickness is large compared with the free path, the deviation in the current density from that which would exist in the bulk metal for the same applied field is appreciable only in a region close to the surface of the specimen. The deviation in the current is proportional to the perimeter P of the cross-section, whilst the mean current is proportional to the cross-sectional area S . Hence

$$\frac{\sigma_0}{\sigma} = 1 + \frac{P}{S} \times \text{constant} = 1 + CIP/S, \quad (35)$$

* Note that Dingle denotes the *radius* of the wire by a .

where C is a dimensionless constant independent of the shape of the specimen. For a thin film of thickness a , $P/S=2/a$, and comparison with (26) shows that $C=3(1-p)/16$. Therefore, for a wire of arbitrary cross-section,

$$\frac{\sigma_0}{\sigma} = 1 + \frac{3}{16} (1-p) \frac{lP}{S} \quad (36)$$

In the notation of § 2.2, $P/S=4/a$ for both square and circular wires, and we regain (32).

TABLE 2

The resistivity of thin wires divided by the resistivity of the bulk metal

κ (diameter/free path)	σ_0/σ	
	$p=0$	$p=\frac{1}{2}$
0.001	1000	337
0.002	503	170
0.005	202	69.3
0.01	102	35.7
0.02	52.1	18.7
0.05	21.6	8.31
0.1	11.45	4.88
0.2	6.33	3.02
0.5	3.14	1.84
1	2.04	1.422
2	1.475	1.208
5	1.172	1.080
10	1.081	1.038
20	1.0390	1.0191
50	1.0152	1.0076
100	1.0076	1.0038

2.32. We may define an 'effective' free path l_{eff} for conduction in thin specimens by writing the effective conductivity in the standard form $\sigma = ne^2 l_{\text{eff}} / m\bar{v}$. For wires of diameter small compared with the free path, we see from equations (33) and (34) that $l_{\text{eff}} = a$, apart from a constant of the order of unity. This result is intuitively obvious; for a very thin film, however, we find that $l_{\text{eff}} \approx a \log(l/a)$, so that l_{eff} tends to infinity with l . The qualitative difference between the behaviour of thin films and of wires may at first sight seem surprising; its origin becomes more apparent if we consider a simple qualitative treatment of conduction in thin films due to Lovell (1936). We assume that all free paths start at the surface, and that the effective free path is the average of all free paths of a given electron, the free path being taken as the distance to the next intersection with the surface or the ordinary free path, whichever is least. Assuming further that the density of electrons travelling inwards from the surface is uniform (diffuse scattering), we have, if θ is the angle between the direction of motion and the normal to the film surface,

$$l_{\text{eff}} = \int_0^{\theta_0} \frac{a}{\cos \theta} \sin \theta \, d\theta + \int_{\theta_0}^{\frac{1}{2}\pi} l \sin \theta \, d\theta,$$

where $\cos \theta_0 = a/l$. This gives

$$l_{\text{eff}} = a \left(\log \frac{l}{a} + 1 \right), \quad \dots \dots \dots (37)$$

which is in qualitative agreement with the exact result when $l \gg a$. Examination of the integrals shows that the main contribution to l_{eff} when $l \gg a$ comes from values of θ close to θ_0 ; the dominant contribution to the current in films, however thin compared with the free path, therefore arises from electrons which, after reflection at the surface of the film, move in directions so nearly parallel to the surface that their free path remains of the order of the bulk free path. In the case of wires the constriction of free paths due to the geometry is more severe, and such electrons do not play any appreciable part.

2.4. *Comparison with experiment.* It was known already at the end of the last century that very thin films of metal exhibit a higher electrical resistivity than the same metal in bulk, and J. J. Thomson (1901) was the first to suggest that the effect arose from the limitation of the free path of the electrons and to give an approximate theory; he derived a formula similar to (37). Since then an enormous amount of experimental work has been carried out on the electrical properties of thin films; we mention in particular only Lovell's careful measurements on films of the alkali metals (Lovell 1936), and refer the reader for further information and references to the critical review by Chambers (1951). However, as we have already emphasized, if the measurements are to give information about the electronic parameters characteristic of the bulk metal, great care must be taken to eliminate all effects which may be peculiar to thin specimens other than the pure geometrical limitation of the free path with which the theory is alone concerned. In particular, both strain and impurity may be expected to play an increasing part in determining the phenomena as the size of the specimen is reduced. In fact it was not until recently that Andrew, working with relatively thick films at liquid-hydrogen and liquid-helium temperatures, obtained data which could be compared with any confidence with the theoretical predictions (Andrew 1949).

For thin wires there is much less experimental work than for films. The most important earlier experiments in this field are those of Eucken and Förster (1934) and succeeding workers, and data on wires were also obtained by Andrew in his recent work. To illustrate the kind of information that can be obtained from the experiments, Andrew's results on both films and wires are compared with the theory below.

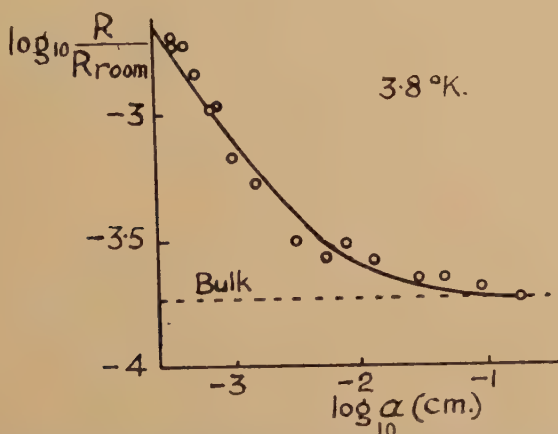
There are two ways in which the ratio a/l may be varied in carrying out the experiments. We may firstly use a single specimen and vary l by changing the temperature. Since the bulk conductivity σ_0 cannot be measured on the specimen itself, it must then be assumed that the temperature variation of σ_0 , and hence of l , is the same for a thin specimen as for a bulk specimen. Alternatively (and this was Andrew's

procedure) σ may be measured at one or more fixed temperatures for a number of specimens of different thickness or diameter; it is then important to ensure that the bulk free path is the same for all specimens.

2.41. Fig. 1 shows Andrew's results on the resistance of rolled foils of tin at 3.8°K . The resistance, expressed as a fraction of the value at 291°K , is plotted logarithmically against the foil thickness, which ranges from 3 to 2000μ .

In comparing these results with the theory, it is necessary to assume a value for the surface reflection coefficient p . Such evidence as exists mostly tends to show that the scattering is diffuse, so that $p=0$ (compare §§ 2.411 and 4.64 below), and this was the value adopted by Andrew; the theoretical curve of σ_0/σ for a film as a function of a/l can then be fitted directly to the experimental results at constant temperature, and the full curve in fig. 1 has been drawn in this way. The rather large

Fig. 1.

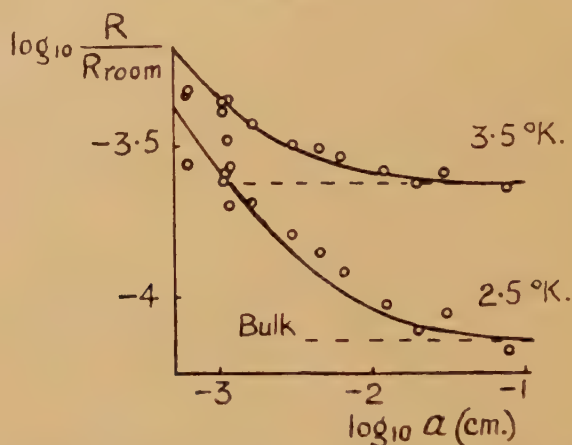


Experimental results on the resistivity of tin foils at 3.8°K , plotted logarithmically against the foil thickness. The full line is the theoretical curve for diffuse scattering ($p=0$).

scatter of the observations about the theoretical curve is probably due to a variation in residual resistivity between the specimens, but on the whole the agreement between theory and experiment is satisfactory. The process of fitting gives directly the value of the free path l corresponding to the temperature of the experiment and the ratio of the bulk conductivity σ_0 to the (known) room-temperature value. The results are that $l=9.5 \times 10^{-3} \text{ cm}$ at 3.8°K , and $\sigma_0/l=4.5 \times 10^{22}$ gaussian units. Measurements at higher temperatures confirm that the value of σ_0/l is constant, and with this value equation (4) gives $n=1.6 \times 10^{22}$. The number of atoms per unit volume, n_a , is 3.7×10^{22} for tin, and thus $n/n_a=0.43$. Such a value for the number of electrons per atom is entirely reasonable, although in a complicated metal such as tin the precise figure has little significance.

2.411. Andrew's measurements on wires were carried out using mercury, and the results for 3.5°K and 2.5°K are shown in fig. 2. The exact analysis of conduction in thin wires had not been given at the time of Andrew's experiments, and the full curves in fig. 2 were obtained on the basis of an approximate theory similar to Lovell's. Dingle (1950) later re-examined Andrew's results and found that the exact theory did not substantially alter the conclusions*. Dingle considered in particular the question of the value to be assigned to p : he found that, if a non-zero value is assumed, agreement with experiment could only be obtained by supposing that p increases slightly with temperature. This is not plausible on theoretical grounds, since the various mechanisms which may lead to diffuse surface scattering, such as the thermal vibration of the surface atoms or the number of mobile surface defects of various kinds, would be

Fig. 2.



Experimental results on the resistivity of mercury wires at 3.5°K and 2.5°K , plotted logarithmically against the wire diameter. The full lines are theoretical curves for diffuse scattering.

expected to become more effective the higher the temperature, causing p to decrease. The value $p=0$ is therefore the most likely; it must be emphasized, however, that more experimental and theoretical work is required before the question can be regarded as definitely settled (compare also § 4.641). Adopting the value $p=0$ and disregarding the results obtained for the thinnest wires used in the experiments, which do not fit in with the theoretical predictions, it is found that $l=5.6 \times 10^{-3} \text{ cm}$ at 2.5°K and $\sigma_0/l=2.5 \times 10^{22}$, which corresponds to 0.15 electrons per atom in mercury.

* It is worth noting that the simple formula $\sigma_0/\sigma=1+1/\kappa$ gives values of σ_0/σ which differ from the exact values (for $p=0$) by less than 5% over the whole range of κ . This formula (or a slight generalization of it) is associated with the name of Nordheim, and was used in the evaluation of the earlier experimental work on thin wires (Nordheim 1934, Eucken and Förster 1934).

2.42. Corresponding experiments have been carried out by MacDonald and Sarginson (1950) on thin wires of very pure sodium; such experiments are of particular interest, since for an alkali metal quantitative agreement with the free-electron theory is to be expected. MacDonald and Sarginson compare their results with the theory developed by them for a square wire, using values of l estimated on the basis of the free-electron theory from the observed conductivity of a bulk sample. ($l \approx 4.5 \times 10^{-3}$ cm at 4.2° K.) The agreement between theory and experiment is only fair, and MacDonald and Sarginson conclude that the hypothesis of entirely diffuse scattering at the boundaries is not generally fulfilled for their specimens, and that the degree of elastic scattering is a function of the specimen size. Further experiments on sodium are considered in §§ 3.2 and 3.3.

2.5. *The 'kinetic theory' solution of thin-conductor problems.* Chambers (1950 a) has recently given an elegant formulation of the theory of thin-conductor problems, in which he writes down the solution of the Boltzmann equation in a form suggested by simple kinetic theory considerations.

In the presence of an electric field E in the x -direction, and with $f=f_0+f_1(\mathbf{v}, \mathbf{r})$, the Boltzmann equation takes the form

$$\mathbf{v} \cdot \text{grad}_{\mathbf{r}} f_1 - \frac{E\epsilon}{m} \frac{\partial f_0}{\partial v_x} = -\frac{f_1}{\tau}. \quad (38)$$

A particular solution is

$$f_1(\mathbf{v}, \mathbf{r}) = \frac{\epsilon\tau E}{m} \frac{\partial f_0}{\partial v_x} \left\{ 1 - \exp\left(-\frac{|\mathbf{r}-\mathbf{r}_0|}{\tau v}\right) \right\}, \quad (39)$$

where $\mathbf{r}-\mathbf{r}_0$ is a vector parallel to \mathbf{v} . This solution applies to conductors of arbitrary shape and, if $\mathbf{r}=\mathbf{r}_0$ is taken to be a point on the surface of the conductor, it satisfies the boundary conditions corresponding to diffuse scattering at the surface.

Equation (39) may be derived from first principles as follows. Consider electrons passing through a point \mathbf{r} in the metal in the direction $\mathbf{r}_0-\mathbf{r}$. The probability that an electron will travel a distance s and then suffer a collision between s and $s+ds$ is $e^{-s/l} ds/l$, but electrons which arrive at \mathbf{r}_0 will certainly collide there if the scattering is diffuse. The mean distance travelled by an electron without collision after passing through \mathbf{r} is therefore

$$l' = \int_0^{|\mathbf{r}-\mathbf{r}_0|} s e^{-s/l} ds/l + \int_{|\mathbf{r}-\mathbf{r}_0|}^{\infty} |\mathbf{r}-\mathbf{r}_0| e^{-s/l} ds/l = l(1 - e^{-|\mathbf{r}-\mathbf{r}_0|/l}); \quad (40)$$

also, for electrons travelling in the opposite direction $\mathbf{r}-\mathbf{r}_0$, the mean distance travelled without collision *before* reaching \mathbf{r} is given by the same expression. The mean drift velocity acquired by these electrons in the electric field is therefore

$$\Delta v_x = \frac{\epsilon E l'}{m v} = \frac{\epsilon \tau E}{m} (1 - e^{-|\mathbf{r}-\mathbf{r}_0|/l}), \quad (41)$$

and f_1 , which is the *change* due to the presence of the field in the number of electrons travelling in the direction $\mathbf{r}-\mathbf{r}_0$ with velocity \mathbf{v} , is given by

$$f_1(\mathbf{v}, \mathbf{r}) = \frac{\partial f_0}{\partial v_x} \Delta v_x = \frac{e\tau E}{m} \frac{\partial f_0}{\partial v_x} (1 - e^{-|\mathbf{r}-\mathbf{r}_0|/l}), \quad (42)$$

in agreement with equation (39).

2.51. The current density may now be calculated exactly as in §2.111, except that the appropriate polar coordinate system (v, δ, ϖ) to be used here has the x -axis as polar axis, this being the only preferred direction which occurs in the present general formulation of the theory. The effective conductivity σ is obtained by integrating the current density over the cross-sectional area S of the conductor, and the ratio of σ to the bulk conductivity σ_0 may be written in the compact form

$$\frac{\sigma}{\sigma_0} = 1 - \frac{3}{4\pi S} \int_S dS \int_0^{2\pi} d\varpi \int_0^\pi d\delta \sin \delta \cos^2 \delta e^{-|\mathbf{r}-\mathbf{r}_0|/l}. \quad (43)$$

To evaluate σ/σ_0 for any shape of conductor it is necessary to express $|\mathbf{r}-\mathbf{r}_0|$ in terms of the position of the point \mathbf{r} and the angles δ and ϖ which define the direction of $\mathbf{r}-\mathbf{r}_0$, and to carry out the integrations. For example, the expressions (15) and (28) given above for a thin film and a thin wire may readily be derived in this way; for a film, however, it is convenient first to change back to the polar coordinates (v, θ, ϕ) used in §2.111. A more complicated problem, which well illustrates the power of the present method, is provided by the case of a thin wire placed in a longitudinal magnetic field; this is discussed in §3.2.

The case of partially elastic scattering has also been considered by Chambers, who derived the appropriate generalization of (43): in this case, however, it is impossible to obtain a simple formula applicable to conductors of arbitrary cross-section.

§3. MAGNETIC EFFECTS IN THIN CONDUCTORS

3.1. In considering conduction in thin specimens when a magnetic field is present, we shall confine attention to the film and the circular wire, and to purely longitudinal or transverse fields. There are then altogether five possible arrangements: these fall into three groups, depending on whether the magnetic field is parallel or perpendicular to the electric current, and whether the Hall electric field which is set up in the latter case is or is not uniform across the specimen.

(A) *H parallel to J*. No Hall field is produced, and this considerably simplifies the analysis. There are two cases in this group:

(A1) the wire with H and J parallel to the axis (§3.2), and

(A2) the film with H and J parallel to each other and in the plane of the film.

(B) *H perpendicular to J: Hall field uniform*. The Hall field is set up in the direction perpendicular to both H and J to prevent any current flow in this direction. If it is developed across a 'thin' dimension of

the specimen it will vary in magnitude across the specimen and its determination as a function of position presents considerable difficulties (see §3.5). In one case, however, namely that of a thin film with H perpendicular to the plane of the film, the Hall field (being in the plane of the film) is developed across a 'long' dimension of the specimen and is consequently constant. This case is in fact very simple to analyse (§3.4).

(C) H perpendicular to J : Hall field non-uniform. The two remaining cases fall into this group: they are

(C1) the film with H in the plane of the film but perpendicular to J , and

(C2) the wire with H perpendicular to the axis.

An approximate treatment has been given of case (C1) (§3.5); case (C2) has not yet been considered, and indeed is of such complexity as to be virtually insoluble.

Exact treatments have so far been given only of cases (A1) and (B), and these will be considered in some detail below. Experimental data are so far available only for wires (cases (A1) and (C2)); the case (A1) of a thin wire in a longitudinal magnetic field is thus of particular interest as being the only case at present for which there exist both an exact evaluation of the theory and experimental results with which to compare it.

A thin-conductor effect of the present type was first observed by MacDonald (1949). The ordinary bulk magneto-resistance effect always leads to an increase in resistance; MacDonald found, however, that the resistance of a thin sodium wire *decreased* with H in a longitudinal field, whilst in a transverse field it increased initially, passed through a maximum and then decreased. Sodium has also been used in subsequent investigations (MacDonald and Sarginson 1950, Chambers 1950 a); this metal (and possibly pure lithium and potassium) appears unfortunately to be the only metal suitable for the experiments in view of the higher inherent bulk magneto-resistance of all other metals, including even the heavier alkali metals (MacDonald 1950). The bulk effect is greatest under just the same conditions of low temperatures and high fields as are required to observe the free electron phenomena, and in general it obliterates the latter entirely.

3.2. *The conductivity of thin wires in a longitudinal magnetic field.* MacDonald, at the time of his discovery of the effect, correctly interpreted its physical origin, and in particular he explained the simple decrease in resistance which occurs in a wire in a longitudinal field as being due to the lessened influence of scattering at the walls of the wire when the electrons are forced to pursue spiral paths around the lines of force of the magnetic field.

An exact analysis of this case (A1) has been given by Chambers (1950 a), assuming diffuse scattering at the boundaries. When the electric and magnetic fields are parallel, the magnetic force on the electrons is always perpendicular to the electric force; we can then regard the electric field

alone as producing a drift current in the usual way, and the magnetic field simply as modifying the electronic trajectories. The solution of the Boltzmann equation is therefore formally the same as in the absence of a magnetic field (equation (42)), and the conductivity is still given by (43), provided that $|\mathbf{r}-\mathbf{r}_0|$ is replaced by the distance s from the point \mathbf{r} to the point on the surface \mathbf{r}_0 measured along the curved trajectory of the electrons.

3.21. The integral (43) was evaluated by Chambers for this case by means of an ingenious series of transformations. With a magnetic field H along the x -axis (the axis of the wire), electrons travelling at an angle δ to the x -axis will move in helical paths whose projections on the yz -plane are circles of radius

$$r = m\bar{v}c \sin \delta / eH = r_0 \sin \delta. \quad (44)^*$$

If now, while an electron is travelling from \mathbf{r}_0 to \mathbf{r} , the projection of its path on the yz -plane traverses an angle ψ around such a circle, then the projection of the distance s on the yz -plane is $\psi r_0 \sin \delta$, so that $s = \psi r_0$. Equation (43) therefore becomes

$$\frac{\sigma}{\sigma_0} = 1 - \frac{3}{4\pi S} \int_S dS \int_0^{2\pi} d\varpi \int_0^\pi d\delta \sin \delta \cos^2 \delta e^{-r_0 \psi / l}, \quad (45)$$

where $\psi = \psi(y, z, \delta, \varpi)$.

3.22. If we consider a fixed value of the azimuth angle ϖ and perform the integrations over S and δ , it is evident by symmetry that the result must be independent of ϖ : we may therefore restrict ourselves to a particular value of ϖ , say $\varpi = 0$. We may furthermore confine ourselves to values of δ in the range $(0, \frac{1}{2}\pi)$, and write (45) in the form

$$\frac{\sigma}{\sigma_0} = 1 - \frac{3}{S} \int_0^{\frac{1}{2}\pi} d\delta \sin \delta \cos^2 \delta \int_S dS (e^{-r_0 \psi / l})_{\varpi=0}. \quad (46)$$

The integrand of (46) refers to electrons which, since colliding with the wall, have turned through angles between ψ and $\psi + d\psi$ and are, at the instant considered, travelling in the direction $(\delta, \varpi = 0)$. If we denote the proportion of the total cross-section of the wire occupied by these electrons by $p(\gamma, \psi) d\psi$ (where γ is defined below), we may rewrite (46) as

$$\frac{\sigma}{\sigma_0} = 1 - 3 \int_0^{\frac{1}{2}\pi} d\delta \sin \delta \cos^2 \delta \int_0^{2\pi} p(\gamma, \psi) e^{-r_0 \psi / l} d\psi. \quad (47)$$

It remains to find an expression for $p(\gamma, \psi)$, which is a function only of ψ and of the ratio of the radius $\frac{1}{2}a$ of the wire† to the radius $r = r_0 \sin \delta$ of the projection of the orbit, i.e. of $\gamma = 2r/a = 2 \sin \delta / \beta$, where $\beta = a/r_0$.

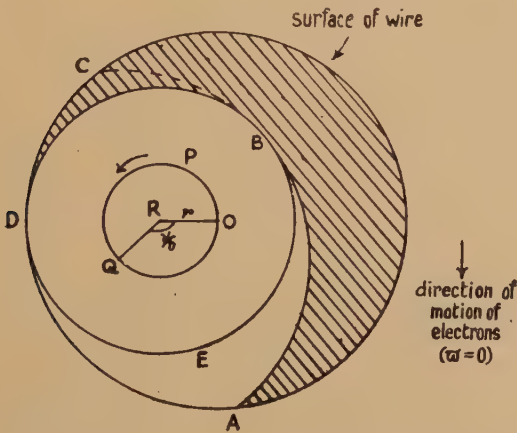
Chambers has given the following construction for the determination of $p(\gamma, \psi)$ (fig. 3). Suppose, for example, that $r \leq \frac{1}{2}a$ ($\gamma \leq 1$), and consider electrons which traverse the projection of their orbit in the anti-clockwise

* r and r_0 in this equation must not, of course, be confused with the magnitudes of the position vectors \mathbf{r} and \mathbf{r}_0 .

† Note that Chambers uses a for the radius of the wire.

direction and which are travelling vertically downwards at the instant considered. Let O be the centre of the wire, and let R be chosen such that OR is horizontal and of length r . With centre R , draw the circles OPQ of radius r and BDE of radius $\frac{1}{2}a - r$. The circle BDE is then such that, for the group of electrons considered, no point within it can be reached by a trajectory starting from the wall. Suppose that OQ subtends an angle ψ_0 at R , and with centre Q and radius $\frac{1}{2}a$ draw the circular arc ABC which touches the circle BDE at B ; it is easily seen that this is the locus of the positions of all electrons of the kind considered which have turned through an angle ψ_0 since colliding with the wall. The portion BC must be excluded, however, since it corresponds to electrons whose trajectory lies partly outside the wire; if, instead, we continue with the arc BD of the circle BDE , then the shaded area between ABD and the surface of the wire contains all those electrons which have turned through angles $\psi \leq \psi_0$. The shaded area is therefore $P(\gamma, \psi_0) = \int_0^{\psi_0} p(\gamma, \psi) d\psi$. $p(\gamma, \psi)$ may be obtained from this by differentiation or, alternatively, (47) may be expressed directly in terms of $P(\gamma, \psi)$ by partial integration.

Fig. 3.


Construction for the determination of $P(\gamma, \psi_0)$ (see text).

Analytical expressions for $p(\gamma, \psi)$ may be obtained by considering the geometry of fig. 3 and the corresponding figure for $\gamma > 1$. It is readily shown that, for $\gamma \leq 1$,

$$\pi p(\gamma, \psi) = \gamma(1 - \gamma) + \gamma^2 \cos^2 \frac{1}{2}\psi + \gamma \cos \frac{1}{2}\psi (1 - \gamma^2 \sin^2 \frac{1}{2}\psi)^{1/2} \quad (0 \leq \psi \leq 2\pi), \quad (48)$$

and, for $\gamma \geq 1$,

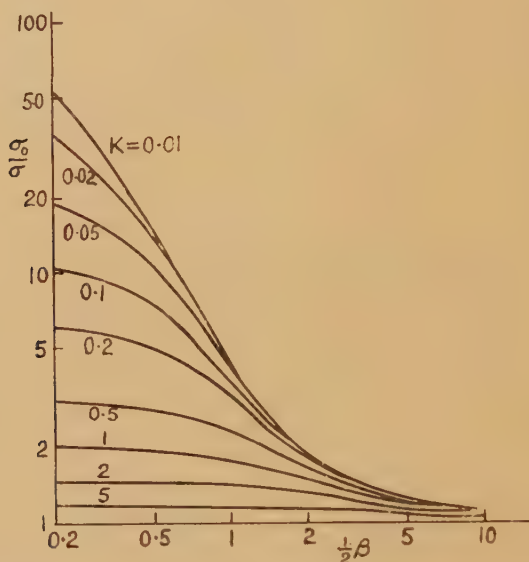
$$\left. \begin{aligned} \pi p(\gamma, \psi) &= 2\gamma \cos \frac{1}{2}\psi (1 - \gamma^2 \sin^2 \frac{1}{2}\psi)^{1/2} & (0 \leq \psi \leq 2 \sin^{-1}(1/\gamma)), \\ &= 0 & (2 \sin^{-1}(1/\gamma) < \psi \leq 2\pi), \end{aligned} \right\} \quad (49)$$

where the positive value of the square root is to be taken.

Note that, for $\beta \geq 2$, $\gamma \leq 1$ for all δ , while for $\beta < 2$, we have $\gamma < 1$ for small δ and $\gamma > 1$ for large δ ($\sim \frac{1}{2}\pi$). Physically this means simply that for strong enough fields ($\beta \geq 2$) all electronic trajectories are curved into paths of radius less than the wire radius; for smaller fields, those electrons travelling at small angles to the axis and therefore having small transverse velocities will still follow such paths, but electrons moving at greater angles to the axis will follow paths of radius greater than the wire radius.

3.23. Values of σ/σ_0 may now be computed by graphical integration using fig. 3, or by approximate analytical methods using equations (48) and (49). The final results, as is clear from (47), depend only on the two dimensionless parameters $\beta = a/r_0$ and $\kappa = a/l$, which are measures of the strength of the magnetic field and of the free path respectively. In

Fig. 4.



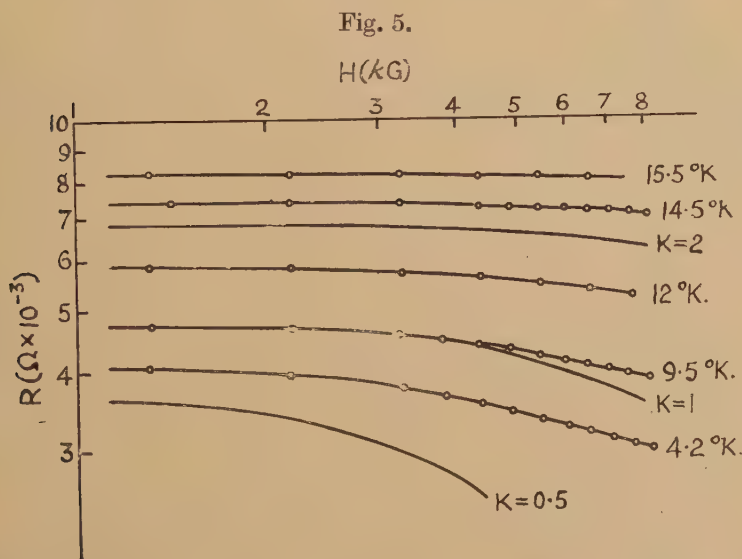
Theoretical conductivity of thin wires in a longitudinal magnetic field.

fig. 4 theoretical values of the resistivity ratio σ_0/σ are shown plotted against $\frac{1}{2}\beta$, for various values of κ . The values for $\beta=0$ are those given by Dingle (§ 2.2), and, as is to be expected on physical grounds, the resistance decreases steadily as β increases and tends to the bulk value as β tends to infinity.

3.24. *Comparison with experiment.* If we plot $\kappa\sigma_0/\sigma$ against $\frac{1}{2}\beta$ for various κ , then, since $\kappa = a/l$ and σ_0/l is independent of the temperature, the ordinates are proportional to the resistivity $1/\sigma$, and, since $\frac{1}{2}\beta = e\hbar H/(2m\hbar v)$, the abscissae are proportional to the applied magnetic field. The theoretical curves of $\kappa\sigma_0/\sigma$ against $\frac{1}{2}\beta$, plotted logarithmically, are therefore directly superposable on the experimental curves of $1/\sigma$ against H at a number of temperatures, and the proportionality constants

give values of σ_0/l and of $m\bar{v}$ directly. The advantage of using a magnetic field is that all the required information can be obtained from one specimen only, and it is not necessary to assume, as in zero-field experiments, that the free path in a thin specimen is the same as in a bulk specimen.

Fig. 5 shows experimental results obtained by Chambers on a sodium wire of diameter 30μ at various temperatures; also shown are the theoretical curves for $\kappa=0.5, 1$ and 2 , adjusted to fit as well as possible. It has, of course, been assumed that the hypothesis of diffuse reflection at the surface is correct, and that the bulk magneto-resistance effect is negligible. The departure from theory in high fields is, in fact, due to the onset of the bulk effect, but this does not interfere seriously with the comparison between theory and experiment. The results of the comparison are that, in sodium, $\sigma_0/l=8.1 \times 10^{22}$ and $m\bar{v}=9.1 \times 10^{-20}$; these values



Experimental results on a 30μ sodium wire in a longitudinal magnetic field at various temperatures, with theoretical curves.

provide two independent estimates of the number of conduction electrons, and agree reasonably well with the values

$$\sigma_0/l=6.3 \times 10^{22} \quad \text{and} \quad m\bar{v}=9.7 \times 10^{-20}$$

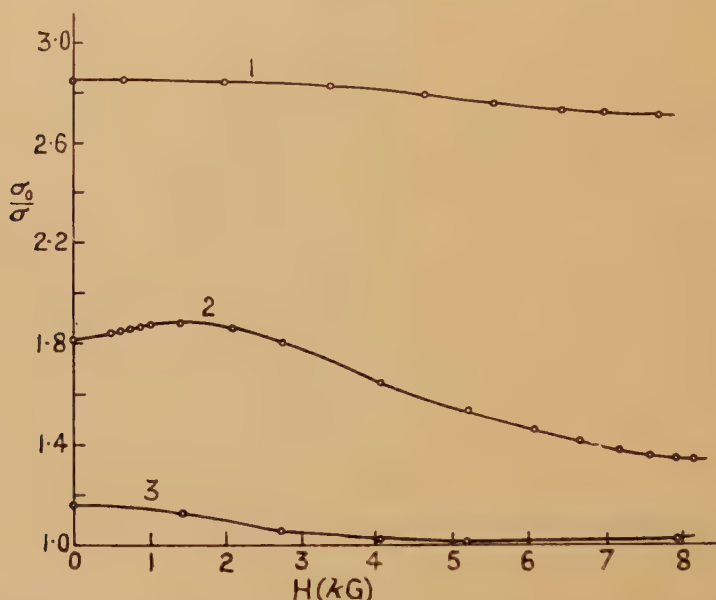
obtained from (4) and (1a) assuming one conduction electron per atom.

3.3. Similar experiments have been reported by MacDonald and Sarginson (1950) on sodium wires in both longitudinal and transverse magnetic fields, and some typical results for transverse fields are shown in fig. 6; note particularly that in this case the resistance may *increase* initially with H . Detailed theoretical analysis of this arrangement (C2) is prohibitively difficult, and MacDonald and Sarginson compared their

results with an approximate theory developed by them for the somewhat simpler thin-film arrangement (C1) (see § 3.5), with which they are in qualitative agreement.

3.4. *The conductivity of thin films in a transverse magnetic field.* The case (B) of a thin film in a magnetic field which is perpendicular to the plane of the film has been analysed by Sondheimer (1950). In this case it is not immediately obvious from simple physical arguments how the presence of the magnetic field affects the conductivity at all. The formal analysis, however, presents no difficulties and is a straightforward generalization of the treatment given in § 2.1 for a film in zero magnetic field: it leads to the interesting result that the conductivity is an oscillatory function of the strength of the applied magnetic field.

Fig. 6.



Experimental results on sodium wires of various diameters in a transverse magnetic field. Diameters: 1, 20 μ ; 2, 30 μ ; 3, 66 μ . Temperature = 4.2° K.

3.41. We use the notation of § 2.1, but suppose now that the film is subjected to an electric field ($E_x, E_y, 0$) in the plane of the film and a magnetic field $(0, 0, H)$. For this particular arrangement, the condition $\text{curl } \mathbf{E} = 0$ ensures that the electric field components are constant across the thickness of the film. The Boltzmann equation takes the form

$$\frac{\partial f_1}{\partial z} + \frac{f_1}{\tau v_z} - \frac{\epsilon H}{mc v_z} \left(v_y \frac{\partial f_1}{\partial v_x} - v_x \frac{\partial f_1}{\partial v_y} \right) = \frac{\epsilon}{m v_z} \left(E_x \frac{\partial f_0}{\partial v_x} + E_y \frac{\partial f_0}{\partial v_y} \right), \quad (50)$$

which is the generalization of equation (9). To solve this equation we put

$$f_1 = (v_x c_1 + v_y c_2) \frac{\partial f_0}{\partial v}, \quad (51)$$

where c_1 and c_2 do not depend explicitly on v_x and v_y . With this form for f_1 , equation (50) leads to two simultaneous equations for c_1 and c_2 ; if we introduce the complex quantities $g=c_1-ic_2$, $\mathcal{F}=E_x-iE_y$, we may write these equations in the compact form

$$\frac{\partial g}{\partial z} + g \left(\frac{1}{\tau v_z} + \frac{i\epsilon H}{mc v_z} \right) = \frac{\epsilon \mathcal{F}}{mv v_z}, \quad (52)$$

and the problem is now formally identical with that of § 2.1.

3.42. The rest of the calculation is entirely analogous to that of § 2.1, and the result may be expressed as follows. We define a complex conductivity σ_c by means of the relation $\bar{J}=\sigma_c \mathcal{F}$, where $\bar{J}=\bar{J}_x-i\bar{J}_y$ is the complex current density averaged in the usual way across the thickness of the film. We further introduce the parameters $\kappa=a/l$, $\beta=a/r_0$ (a being the thickness of the film), and the complex variable $s=\kappa+i\beta$. We then have, as the generalization of (16),

$$\frac{\sigma_0}{\sigma_c} = \frac{\Phi(s)}{\kappa}, \quad (53)$$

where the function Φ is given as before by (17) when $p=0$ and by (25) when $p \neq 0$, but is now a function of a complex variable. (We may remark here that the present problem is the only one of the magnetic-field problems which can easily be solved for the case $p \neq 0$.)

3.43. Equation (53) contains all the results required for comparison with experiment. The ordinary electrical conductivity σ , for example, is measured by applying an electric field in, say, the x -direction and observing the current in this direction, no electric current being allowed to flow in the transverse direction. We therefore have $\sigma=\mathcal{R}(\bar{J})/\mathcal{R}(\mathcal{F})$, with $\mathcal{I}(\bar{J})=0$, where \mathcal{R} and \mathcal{I} denote the real and imaginary parts respectively; eliminating the transverse electric field, we readily find that

$$\frac{\sigma_0}{\sigma} = \sigma_0 \mathcal{R} \left(\frac{1}{\sigma_c} \right) = \mathcal{R} \{ \Phi(s) \} / \kappa. \quad (54)$$

The Hall coefficient is defined by $A_H=E_y/H\bar{J}_x$, where E_y is the transverse electric field set up under the above experimental conditions, and we easily obtain

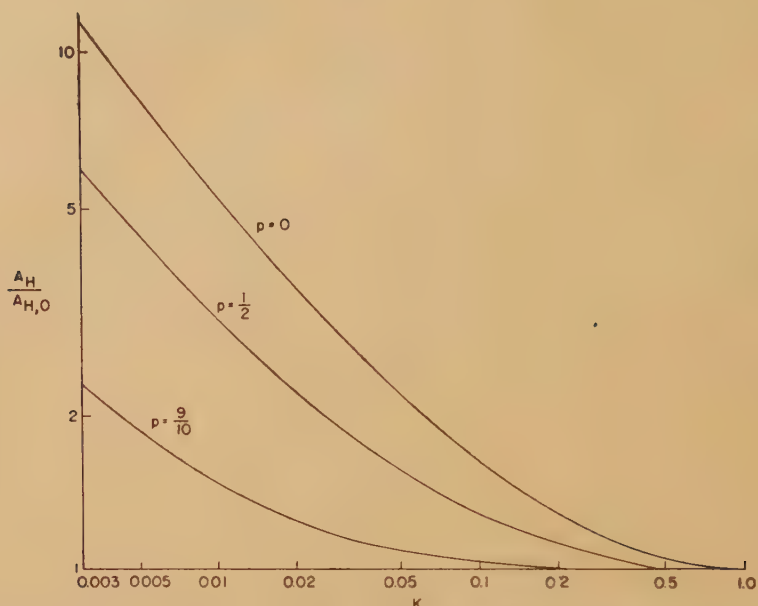
$$A_H/A_{H,0}=\mathcal{I} \{ \Phi(s) \} / \beta, \quad (55)$$

where $A_{H,0}=-1/(n\epsilon c)$ is the Hall coefficient of the bulk metal.

3.44. In the limit of zero magnetic field ($\beta=0$), equation (54) for the conductivity reduces to equation (16), and (55) gives a corresponding expression for the Hall coefficient of a thin film in a vanishingly small magnetic field. In this limit $A_H/A_{H,0}$ depends only on κ , and is shown graphically in fig. 7 for three values of p ; it is seen that the Hall coefficient of a thin film shows an increase above the bulk value analogous to the increase of the electrical resistivity. For very small κ , we find that

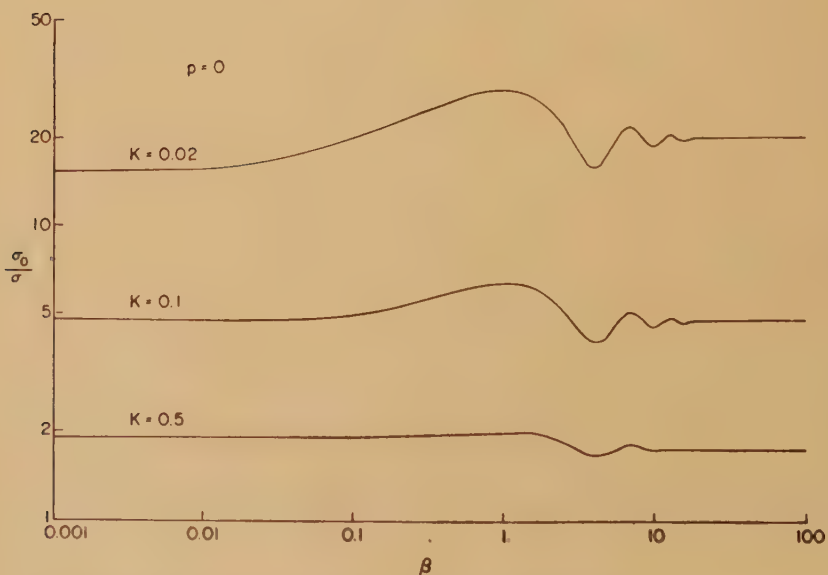
$$\frac{A_H}{A_{H,0}} = \frac{4}{3} \frac{1-p}{1+p} \frac{1}{\kappa \{ \log(1/\kappa) \}^2}, \quad (56)$$

Fig. 7.



Theoretical Hall coefficient of thin metallic films (in small magnetic fields) divided by the Hall coefficient of the bulk metal.

Fig. 8.



Theoretical variation of resistance of thin metallic films with magnetic field perpendicular to surface of film, assuming diffuse surface scattering.

which shows that the Hall coefficient increases less rapidly than the resistance (compare (27)).

3.45. The field variation of the resistivity may be computed from (54), and, as mentioned above, it is found that the resistance oscillates with the strength of the applied magnetic field. Fig. 8 shows some typical curves for $p=0$ and various values of κ . In general the resistance increases initially with H and reaches its first maximum when β is approximately unity; the higher oscillations are of roughly constant spacing but decrease rapidly in amplitude, and in very strong fields the resistance tends to the constant asymptotic value

$$\left(\frac{\sigma_0}{\sigma}\right)_{\beta=\infty} = 1 + \frac{3}{8\kappa}(1-p), \quad (57)$$

which, for small κ , is greater than the value in zero field. The oscillations die out and the resistance tends towards the bulk value as κ becomes large or as p tends to unity; the *position* of the maxima and minima, however, is almost independent of κ and p , and β may thus be estimated unambiguously by observing the value of H for which, for example, the first maximum is reached (assuming, however, as usual that the bulk magneto-resistance effect is negligible).

The Hall coefficient, on the other hand, does not oscillate with H , and decreases steadily towards the bulk value as H increases to infinity.

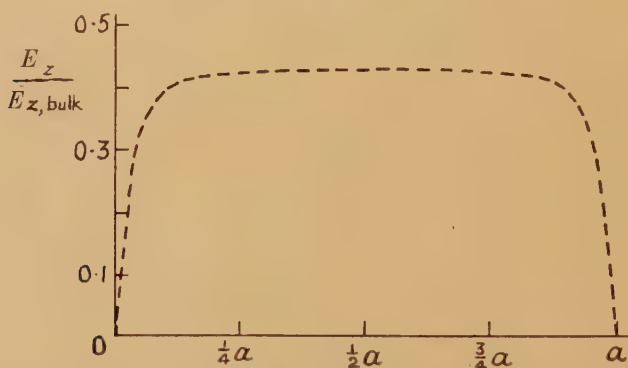
3.46. The resistance oscillations have been explained by Chambers (1950 a) as being essentially due to the oscillations in speed of an electron moving in perpendicular electric and magnetic fields. The speed at time t is a trigonometric function of $eHt/mc = \bar{v}t/r_0$, and this causes the distribution function $f_1(\mathbf{v}, z)$ to be a fluctuating function of z/r_0 . The effect occurs only in thin films because the presence of the metal surface is required to provide a finite limit to the distance from which electrons can come to contribute to the current at z ; when the thickness is large compared with the free path, the elementary oscillating contributions to $f_1(\mathbf{v}, z)$ add up to give a non-fluctuating total, and in fact f_1 is then independent of z .

Equation (57) shows that, in a strong magnetic field, the effective free path of the electrons (see § 2.32) is of the order of the thickness of the film. We therefore have the physically plausible result that a strong transverse magnetic field eliminates the contribution to the current of those electrons moving nearly parallel to the surface which are responsible in the absence of a field for the logarithmic dependence of l_{eff} on the ratio l/a .

3.5. MacDonald and Sarginson (1950) have given an elaborate discussion of a thin film for the case (C1) where the magnetic field is in the plane of the film but perpendicular to the current. For the details of the analysis the reader is referred to the original paper. The Hall field is now in the z -direction (perpendicular to the plane of the film) and varies across the thickness of the film. It can, in principle, be evaluated as a

function of z and H by solving a complicated integral equation which expresses the condition that the current density in the z -direction must vanish everywhere. MacDonald and Sarginson have evaluated the Hall field approximately for the case of small magnetic fields, and have found that it increases rapidly near the edges of the film but is approximately constant over most of the cross-section, as shown in fig. 9. In their calculation of the conductivity MacDonald and Sarginson were forced, however, to assume for simplicity that the Hall field was constant over the whole cross-section. In view of this assumption the results must be treated with caution: they indicate that there are no oscillations, but that, for sufficiently thin films, the resistance increases initially with H , passes through a single maximum and then decreases steadily towards the bulk metal value.

Fig. 9.



Theoretical variation of Hall field E_z across a thin film of thickness $a=0.64 \times$ free path, assuming diffuse scattering at the boundaries.

3.6. The only magnetic-field experiments on thin films so far reported (MacDonald and Sarginson 1950) have been carried out on foils of silver, gold and tin. Although small indications of the effects discussed above were observed, it was found that the bulk magneto-resistance effect was the dominating factor in all cases, and no definite conclusions could be drawn. Experiments on thin films of sodium in magnetic fields both in and perpendicular to the plane of the film are highly desirable in order to test the theoretical predictions; at the same time the theory should be rounded off by a discussion of the third case (A2) where the magnetic field is in the plane of the film and parallel to the current.

§ 4. THE ANOMALOUS SKIN EFFECT

4.1. H. London (1940) discovered that the resistivity of metals at high frequencies and very low temperatures was much greater than is predicted by the usual theory of the skin effect. He suggested that a free-path phenomenon was involved, but did not elaborate his idea. The

problem was later taken up by Pippard (1947), who studied it in detail both experimentally and theoretically; using mainly physical arguments he succeeded in elucidating all the essential features involved. The quantitative theory of the effect is based on the methods discussed in the present article and is due to Reuter and Sondheimer (1948).

4.2. *The normal skin effect.* The basic concepts are best introduced by considering the theory of the skin effect in metals in its simplest form. Let the surface of a semi-infinite metal be the xy -plane and let the positive z -axis be directed towards the interior of the metal. The electric field $E(z)e^{i\omega t}$ is taken to be in the x -direction, ω being the angular frequency, and the magnetic field $H(z)e^{i\omega t}$ is in the y -direction; the factor $e^{i\omega t}$ which occurs in all time-dependent quantities need not be written down explicitly and will in future be omitted.

Neglecting the displacement current, Maxwell's equations take the form

$$-H'(z) = 4\pi J(z)/c, \quad E'(z) = -i\omega H(z)/c, \quad . \quad . \quad . \quad (58)$$

where $J(z)$ is the current density. Eliminating H , we obtain

$$E''(z) = 4\pi i\omega J(z)/c^2, \quad . \quad . \quad . \quad . \quad . \quad (59)$$

and, combining this with Ohm's law in the form $J = \sigma_0 E$, we have

$$E''(z) = 4\pi i\omega\sigma_0 E(z)/c^2, \quad . \quad . \quad . \quad . \quad . \quad (60)$$

so that

$$E(z) = E(0)e^{-kz}, \text{ with } k = (1+i)\sqrt{(2\pi\omega\sigma_0)/c} = (1+i)/\delta, \quad . \quad (61)$$

$$\delta = c/\sqrt{(2\pi\omega\sigma_0)} \quad . \quad . \quad . \quad . \quad . \quad (62)$$

being the depth of penetration of the field (the 'classical skin depth').

In high-frequency experiments the quantity most directly accessible to measurement is not the resistivity of the metal in the ordinary sense, but the *surface* resistivity R , which is the real part of the surface impedance Z . Z is defined as the ratio of the electric field at the surface of the metal to the total current per unit area of surface, and thus

$$Z = R + iX = E(0) \left/ \int_0^\infty J(z) dz \right., \quad . \quad . \quad . \quad . \quad . \quad (63)$$

X being the surface reactance. Equations (59) and (61) give

$$R = X = \sqrt{(2\pi\omega/c^2\sigma_0)}, \quad . \quad . \quad . \quad . \quad . \quad (64)$$

and R and X therefore vary directly as $\sqrt{\omega}$ for a specimen at a given temperature, and inversely as $\sqrt{\sigma_0}$ for a given frequency.

4.21. These formulae are in good agreement with observation on metals at room temperature for all wavelengths in the radio-frequency region. Pippard found, however, that as σ_0 is increased by lowering the temperature, the surface conductivity $1/R$ increases less rapidly than is predicted by equation (64) and that, as σ_0 tends to infinity, $1/R$ tends to a constant value which is different for each metal; the typical form of the curves is shown in fig. 11 below. This is the phenomenon known as the anomalous skin effect, and it evidently involves a breakdown of the simple theory outlined above.

4.22. In the above form the theory is essentially phenomenological in character and makes no direct reference to electron theory; it is based only on Maxwell's equations and the assumption that Ohm's law may be used to relate the current at a point in the metal to the electric field at the same point. It is well known that Ohm's law must be generalized by introducing a complex conductivity when the frequency is so high that the period of the applied field becomes comparable with the time of relaxation of the electrons (Drude 1904, compare also Wilson 1936, p. 124). Relaxation effects of this nature, however, play no significant role in the anomalous skin effect except at infra-red frequencies; they are considered briefly in §4.71, but may be disregarded for the present. The generalization of Ohm's law which concerns us here is of a different type, and is connected with the very long free path of the conduction electrons when the conductivity is high; when the free path is comparable with the skin depth δ , it is not permissible to assume that an electron moves under the influence of a constant field between collisions, and the current at a point in the metal will depend on values of the electric field at other points.

Mathematically, the departure from Ohm's law is a consequence of the appearance of the usual term $\mathbf{v} \cdot \text{grad}_r f$ in the Boltzmann equation. The present problem differs from most of the previous ones, however, in that the electric field in the metal is not given *a priori*, but must itself be determined by the theory. When the relation between the current and an arbitrary field $E(z)$ has been found by solving the Boltzmann equation, it must be combined with Maxwell's equation (59) to give an equation from which $E(z)$ can be calculated; the surface impedance can then be obtained from the equation

$$Z = - \frac{4\pi i \omega}{c^2} \frac{E(0)}{E'(0)}, \quad \dots \dots \dots (65)$$

which follows directly from (59) and (63).

4.3. *The fundamental equation.* So long as we are only interested in the surface impedance of the metal (and not, for example, in the transmission coefficient of a thin film), the penetration depth of the field may always be treated as small compared with the linear dimensions of the specimen, and we may continue to consider a semi-infinite slab of metal. The Boltzmann equation then takes the same form (9) as in the discussion of the d.c. conductivity of a thin film, namely

$$\frac{\partial f_1}{\partial z} + \frac{f_1}{\tau v_z} = \frac{\epsilon}{m v_z} \frac{\partial f_0}{\partial v_x} E(z), \quad \dots \dots \dots (66)$$

where the electric field is now, however, a function of z . (The influence of the high-frequency magnetic field may be neglected in writing down equation (66).) The general solution is now

$$f_1(\mathbf{v}, z) = \exp\left(-\frac{z}{\tau v_z}\right) \left\{ F(\mathbf{v}) + \frac{\epsilon}{m v_z} \frac{\partial f_0}{\partial v_x} \int^z E(t) \exp\left(\frac{t}{\tau v_z}\right) dt \right\}. \quad (67)$$

The value of $F(\mathbf{v})$ when $v_z > 0$ is determined in the usual way by the nature of the scattering at the surface of the metal, which is described by a reflection coefficient p ; when $v_z < 0$, $F(\mathbf{v})$ must be chosen so that f_1 does not become exponentially large as $z \rightarrow \infty$. Defining $E(-z) = E(z)$ for $z < 0$, the solution which satisfies the boundary conditions may be written in the form

$$\left. \begin{aligned} f_1^+(\mathbf{v}, z) &= \frac{\epsilon}{mv_z} \frac{\partial f_0}{\partial v_x} \exp\left(-\frac{z}{\tau v_z}\right) \left\{ p \int_{-\infty}^z E(t) \exp\left(\frac{t}{\tau v_z}\right) dt \right. \\ &\quad \left. + (1-p) \int_0^z E(t) \exp\left(\frac{t}{\tau v_z}\right) dt \right\} \quad (v_z > 0), \\ f_1^-(\mathbf{v}, z) &= -\frac{\epsilon}{mv_z} \frac{\partial f_0}{\partial v_x} \exp\left(-\frac{z}{\tau v_z}\right) \int_z^{\infty} E(t) \exp\left(\frac{t}{\tau v_z}\right) dt \quad (v_z < 0). \end{aligned} \right\} \quad (68)$$

The first of these expressions shows that the effect of partially elastic scattering may be described formally by replacing the semi-infinite metal slab by an infinite metal in which the field in the region $z < 0$ is an image of the field in the region $z > 0$ and in which a fraction p of the electrons moves through the image field.

4.31. The calculation of the current density is now straightforward, and the result is

$$J(z) = \frac{3\sigma_0}{4l} \left\{ p \int_{-\infty}^{\infty} k\left(\frac{z-t}{l}\right) E(t) dt + (1-p) \int_0^{\infty} k\left(\frac{z-t}{l}\right) E(t) dt \right\}, \quad (69)$$

where

$$k(u) = \int_1^{\infty} \left(\frac{1}{s} - \frac{1}{s^3} \right) e^{-s|u|} ds; \quad (70)$$

it is seen that the current density at z involves the values of the electric field at *all* points in the metal. The two integrals on the right-hand side of (69) represent the contributions to J of the electrons which suffer specular and diffuse reflection respectively at the surface of the metal. We may note here that in the anomalous skin effect the value of p plays only a subordinate part in determining the surface impedance (see § 4.61); there is thus a basic difference between the present phenomenon and the d.c. thin-conductor free-path effects which are in general entirely due to the presence of diffuse scattering at the boundaries.

4.32. The equation for E is obtained by combining (59) and (69). It is convenient to introduce dimensionless coordinates $x = z/l$, $y = t/l$, and to write $E(lx) = f(x)$; the fundamental equation of the problem is then obtained in the form

$$f''(x) = i\alpha \left\{ p \int_{-\infty}^{\infty} k(x-y)f(y) dy + (1-p) \int_0^{\infty} k(x-y)f(y) dy \right\}, \quad (71)$$

where

$$\alpha = 3\pi\omega l^2 \sigma_0 / c^2 = \frac{3}{2} l^2 / \delta^2. \quad (72)$$

For the evaluation of the surface impedance it is only necessary to know $f(0)/f'(0)$ (see equation (65)), and $f(x)$ is not required explicitly.

It is easily shown that the results of § 4.2 are regained in the limit where $\alpha \ll 1$.

4.4. *Solution of the fundamental equation.* In the two limiting cases $p=1$ and $p=0$ equation (71) can be solved by standard methods based on the theory of Fourier integrals (Titchmarsh 1937). When $p=1$, for example, the equation reduces to

$$f''(x) = i\alpha \int_{-\infty}^{\infty} k(x-y)f(y) dy, \quad (73)$$

where $f(x)$ is, by definition, an even function of x . It is easy to see that $f'(x)$ is discontinuous at $x=0$, and tends to a limit μ (say) as $x \rightarrow +0$ ($-\mu$ as $x \rightarrow -0$). The formal solution of (73) now proceeds as follows. Let

$$\phi(t) = \int_{-\infty}^{\infty} f(x)e^{-ixt} dx = 2 \int_0^{\infty} f(x) \cos xt dx \quad . . . (74)$$

and

$$\kappa(t) = \int_{-\infty}^{\infty} k(x)e^{-ixt} dx. \quad (75)$$

From (74), integrating by parts twice, it follows that

$$\int_{-\infty}^{\infty} f''(x)e^{-ixt} dx = -t^2\phi(t) - 2\mu; \quad (76)$$

on the other hand, (73) implies that

$$\begin{aligned} \int_{-\infty}^{\infty} f''(x)e^{-ixt} dx &= i\alpha \int_{-\infty}^{\infty} e^{-ixt} dx \int_{-\infty}^{\infty} k(x-y)f(y) dy \\ &= i\alpha\kappa(t)\phi(t) \quad (77) \end{aligned}$$

(inverting the order of integration). Hence, from (76) and (77),

$$\phi(t)\{t^2 + i\alpha\kappa(t)\} = -2\mu,$$

and the inversion formula for Fourier transforms gives the final result

$$f(x) = -\frac{2\mu}{\pi} \int_0^{\infty} \frac{\cos xt dt}{t^2 + i\alpha\kappa(t)}. \quad (78)$$

For $x=0$, in particular, it follows that

$$\frac{f(0)}{f'(0)} = -\frac{2}{\pi} \int_0^{\infty} \frac{dt}{t^2 + i\alpha\kappa(t)}. \quad (79)$$

The explicit form of $\kappa(t)$ is readily obtained from (70) and (75), and is

$$\kappa(t) = 2t^{-3}\{(1+t^2)\tan^{-1}t - t\}. \quad (80)$$

4.41. When $p=0$, equation (71) is of the same general type as the well-known integral equation of Milne which has been much studied in the theory of radiative transfer and in neutron diffusion problems. The solution in this case is based on the method which has been applied to Milne's equation by Wiener and Hopf; for details of the analysis the reader is referred to Reuter and Sondheim's paper. The results are complicated but lead to conclusions which are generally similar to those found for $p=1$ (see § 4.6).

4.5. *The electric field.* The electric field $f(x)$ is not itself required for the evaluation of the surface impedance, but its behaviour is of interest. It is clear from (78) that $f(x)$ is not of exponential form, and in general its evaluation is complicated; for large values of x , however, it may be shown (Reuter and Sondheimer 1948, appendix III) that, when $p=1$,

$$f(x) = C_1 e^{-s_1 x} + C_2 e^{-x/x^2}, \quad \dots \quad (81)$$

where C_1 and C_2 are constants, and where s_1 is a complex parameter whose real part $\Re s_1$ increases steadily with α and equals unity when $\alpha = \alpha_0 \approx 2.63$.

The physical significance of this result is best understood by considering the case of a metal which is unbounded in all directions (Pippard, Reuter and Sondheimer 1948). In this case, when $\alpha \leq \alpha_0$, the electric field is found to be the simple damped exponential wave $e^{-s_1 x}$, the penetration depth of which is $l/\Re s_1$ and is thus equal to l when $\alpha = \alpha_0^*$. For $\alpha > \alpha_0$ there are no solutions at all, which means that an exponential wave which is attenuated by a factor e in a distance smaller than the free path cannot exist: it would correspond to a situation in which the contribution to the current at a point is greater the further the electrons contributing to the current have travelled in the wave direction, so that the expression for the resultant current density at any point diverges.

It is therefore clear that, in order to obtain solutions for all values of α , it is essential to take the boundary conditions at the surface of the metal specifically into account in the calculation; it is to be expected, furthermore, that there will be a fundamental difference in the nature of the electric field for values of α less or greater than the critical value α_0 . This is in accordance with the form of equation (81). When $\alpha < \alpha_0$, the first term is dominant and the field in the metal, at sufficient distances from the surface, is unaffected by the presence of the surface, being similar in form to the solution for an infinite metal; when $\alpha > \alpha_0$, however, the second term is dominant, and the form of the excitation is now conditioned at all points by the presence of the surface and never approximates to a simple exponential form, being, in fact, a true surface excitation. In the latter case the electric field, though largely confined to the surface, has a long 'tail' of small amplitude which is effectively damped out in a distance of the order of l and which may be regarded as being transmitted into the metal by electrons which move in the interior under the influence of no forces except their collisions with the lattice.

4.6. *The surface impedance.* The surface impedance can be obtained from (65) and (79) (or the corresponding expression for $p=0$), and in general the integrals have to be evaluated numerically. Explicit expressions for Z can, however, be obtained in the limiting case ($\alpha \gg 1$) when the free path is very large.

4.61. *The asymptotic expressions for Z .* The limiting value of $f(0)/f'(0)$ when $\alpha \gg 1$ is obtained from (79) by replacing $\kappa(t)$ by its approximate

* We recall that x is the distance measured in units of l .

value π/t for large t . The integral is then readily evaluated and leads to the following limiting value of Z when $p=1$:

$$Z_{\infty} = \frac{8}{9}(\sqrt{3}\pi\omega^2 l/c^4\sigma_0)^{1/3}(1+\sqrt{3}i). \quad (82)$$

The corresponding result when $p=0$ is

$$Z_{\infty} = (\sqrt{3}\pi\omega^2 l/c^4\sigma_0)^{1/3}(1+\sqrt{3}i). \quad (83)$$

These expressions are to be contrasted with the result (64) which holds when $\alpha \ll 1$. We see that Z is independent of l when the free path is very large, in accordance with the experimental results; further, Z varies with frequency as $\omega^{2/3}$, and the surface resistivity and reactance are related by

$$X_s = \sqrt{3}R_s \quad (84)$$

under extreme anomalous conditions. These results hold both for $p=1$ and for $p=0$, and it is evident that although the precise nature of the surface scattering plays some part in determining the surface impedance, the effect is not considerable.

4.62. *The 'ineffectiveness concept'.* The physical principles underlying these results are not readily apparent from the detailed calculation, and it is therefore worth while to outline a simple qualitative theory which was developed by Pippard (1947) before the exact solution had been obtained. When the free path is large compared with the penetration depth, the electrons may be divided into two groups, those moving at small angles to the surface which have a reasonable chance of colliding in the surface layer, and the rest whose chance of such a collision is small. Pippard gave reasons for supposing that only the former group of electrons contributes effectively to the resistance, and the proportion of effective electrons may thus be written $\beta\delta'/l$, where δ' is the (unknown) effective penetration depth and where β is a numerical constant of order unity. It is now supposed that the ineffective electrons may be entirely neglected, and that the effective electrons move in a constant field during a free path, so that the theory of the normal skin effect applies to them. The effective conductivity is then $\sigma' = \beta\delta'\sigma_0/l$, and δ' is given by an expression of the form (62) with σ_0 replaced by σ' . We thus obtain an implicit equation for δ' , which, when solved, gives

$$\delta' = (c^2 l / 2\pi\omega\beta\sigma_0)^{1/3}. \quad (85)$$

Inserting σ' instead of σ_0 in (64) and using (85), we obtain the surface resistivity in the form

$$R_s = (4\pi^2\omega^2 l \beta c^4 \sigma_0)^{1/3}, \quad (86)$$

in qualitative agreement with (82) and (83).

Later work on the interpretation of the theory (Marcus, to be published) has shown that the physical picture underlying the ineffectiveness concept must not be taken too literally, but it does provide a useful guiding principle which is of particular value when considering more complicated situations for which no exact theory exists (compare, for example, § 4.72).

4.63. Since $E(0)/E'(0)=lf(0)/f'(0)$, equations (65) and (79) show that Z depends on the free path both through l and α . It is convenient to express Z as a function only of α and of quantities independent of l ; using equations (1b) and (72) to eliminate l , we find that

$$Z=-i\sqrt{\left(\frac{8}{3}\right)}A\alpha^{1/3}f(0)/f'(0), \quad . \quad . \quad . \quad . \quad . \quad (87)$$

where

$$A=\sqrt{6}\left(\frac{\pi\omega}{\epsilon c^2}\right)^{2/3}\left(\frac{m\bar{v}}{3n}\right)^{1/3} \quad . \quad . \quad . \quad . \quad . \quad (88)$$

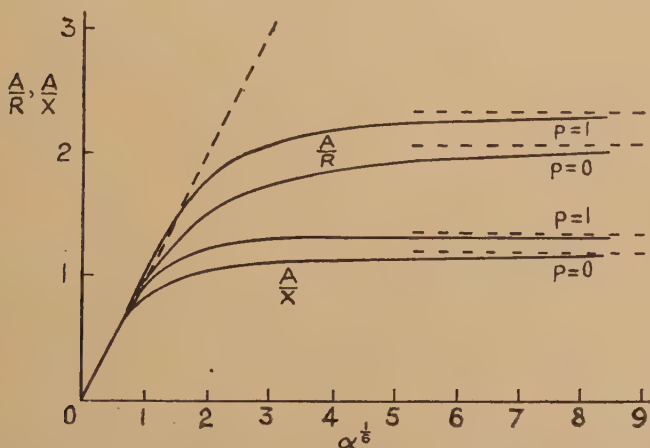
A is constant for a given metal and a given frequency and has been chosen such that, when $\alpha \ll 1$,

$$Z=A\alpha^{-1/6}(1+i), \quad . \quad . \quad . \quad . \quad . \quad (89)$$

as is easily verified. When $\alpha \gg 1$, we obtain

$$Z_\infty=\frac{4\sqrt{2}}{9\pi^{1/3}}A(1+\sqrt{3}i) \quad (p=1) \quad . \quad . \quad . \quad (90)$$

Fig. 10.



The theoretical high-frequency surface impedance of metals at low temperatures.

and

$$Z_\infty=\frac{A}{\sqrt{2\pi^{1/3}}}(1+\sqrt{3}i) \quad (p=0); \quad . \quad . \quad . \quad (91)$$

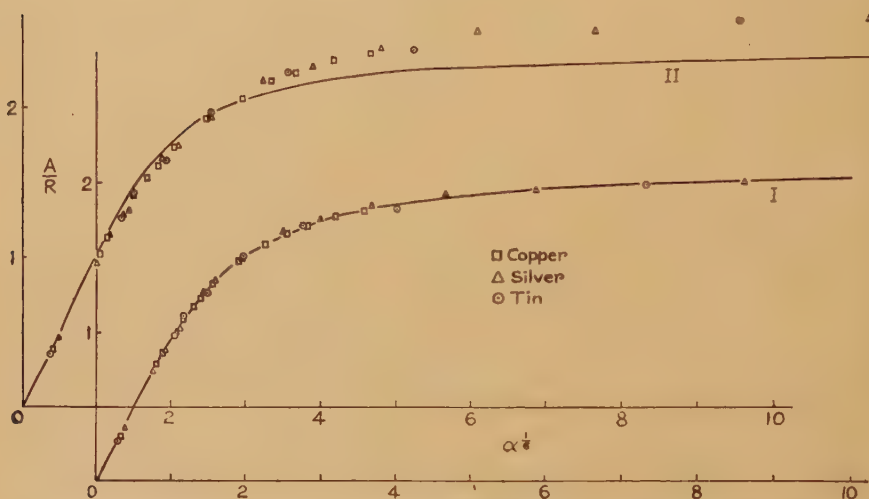
equations (89), (90) and (91) are of course equivalent to (64), (82) and (83). The complete theoretical curves are shown in fig. 10, where A/R and A/X have been plotted against $\alpha^{1/6}$.

4.64. *Comparison with experiment.* The experimental method used by Pippard to measure R involves the construction of a resonator of the metal to be studied and the determination of its selectivity Q ; for a given resonator, Q is proportional to the surface conductivity $1/R$. The results for any metal are exhibited by plotting $1/R$ against $\sqrt{\sigma_0}$ at constant frequency. Since $\sqrt{\sigma_0}$ and $\alpha^{1/6}$ both vary as \sqrt{l} , the experimental points

can be scaled directly to fit the theoretical curve if it is assumed that p equals either 1 or 0. It is then possible, by reading off the value of α corresponding to any pair of values of R and σ_0 , to estimate σ_0/l for any metal: according to equations (62) and (72), $\sigma_0/l = \sqrt{(3\pi\omega\sigma_0^3/c^2\alpha)}$.

Pippard's (1947) results were in qualitative agreement with the theory, but were not sufficiently detailed to indicate which value is to be chosen for p . Chambers (1950 b) has recently carried out more extensive experiments, in which the temperature was varied between 2°K and 90°K ; the frequency (as in Pippard's experiments) was 1200 Mc/s, corresponding to a free-space wavelength of 25 cm. Some of Chambers' results for copper, silver and tin are shown in fig. 11, scaled (I) to fit the theoretical curve for $p=0$, and (II) in an attempt to fit the curve for $p=1$.

Fig. 11.



Experimental results on the surface resistance at low temperatures.

(I) Experimental points fitted to $p=0$ curve.

(II) Attempt to fit to $p=1$ curve.

The agreement with curve I is really excellent, and it thus appears that the reflection at the surface is completely diffuse at all temperatures for all specimens.

4.641. These experiments, as Chambers has pointed out, suggest strongly that the value of p which describes the situation in real metals is zero in all cases, including the d.c. size effects (compare §§ 2.41 and 2.411). It should be noted that the appropriate value of p need not necessarily be the same for all the different problems studied: although p has been defined in the same way in each case, it probably represents some kind of average value of some more complicated quantity, and there is no reason to suppose that this average should have the same value in all cases. If Pippard's ineffectiveness concept is taken literally, the

value of p in the anomalous skin effect is determined by the behaviour on reflection of electrons travelling at very small angles to the surface. If specular reflection occurs at all, it seems most likely to occur for these electrons, and since it is not observed for them (i.e. since $p=0$ in the anomalous skin effect), it is reasonable to expect that $p=0$ also for the d.c. case, where electrons travelling at all angles to the surface are 'effective'.

4.642. Values of σ_0/l deduced from the experiments are shown in table 3, together with the corresponding values of n/n_a , the number of conduction electrons per atom. The values for copper, silver, gold and tin were obtained by Chambers*, and are subject to probable errors of about $\pm 5\%$; the values for mercury and aluminium have been deduced from Pippard's measurements and have larger probable errors.

TABLE 3.

Estimates of σ_0/l and n/n_a from observations on the anomalous skin effect.

Metal	$(\sigma_0/l) \times 10^{-22}$ (gaussian units)	n/n_a
Cu	13.9	1.0
Ag	8.3	0.68
Au	7.6	0.60
Sn	8.6	1.1
Hg	3.3	0.23
Al	5.5	0.37

The results for the monovalent metals are in fair agreement with the expected value of one conduction electron per atom, although the figures for silver and gold are somewhat too low. The figures for tin and mercury may be compared with the corresponding figures ($n/n_a=0.43$ and 0.15) obtained from the thin-conductor experiments: the agreement is probably as good as can be expected for such complicated metals.

4.643. The present method of estimating the free path is free from the main difficulties which beset the thin-conductor experiments, such as the impossibility of measuring the bulk conductivity directly on the specimen, or the disturbing influence of the bulk magneto-resistance effect. It is, however, very important to ensure that the specimens have no surface layer with properties different from those of the bulk metal: if, for example, there is a thin surface layer, of thickness comparable with the skin depth, with an abnormally high residual resistance, the temperature variation of the surface resistivity will simulate the behaviour in the anomalous skin effect, although in this case the high resistivity will not be a result of the long free path. This is in fact what appears to have happened in early measurements by Pippard on wires of the noble metals, which led to very low values of n/n_a . The process of drawing wires of

* These values differ from those given in the published note (Chambers 1950 b); they are based on more accurate values of the d.c. conductivity σ_0 .

these metals produces an amorphous layer which is not easily removed by annealing, but which, in Chambers' experiments, was removed by electrolytic polishing. In subsidiary experiments Chambers has confirmed that the predominant cause of abnormally high surface losses is the presence of a high-resistance layer, and that possible alternative causes such as the roughness of the surface are generally of smaller importance.

Similar disturbing influences appear to have played a part in experiments on the anomalous skin effect carried out in the United States (Fairbank 1949, Maxwell, Marcus and Slater 1949).

4.644. Pippard (1950 a) has recently made experiments on tin using a frequency of 9400 Mc/s, corresponding to a wavelength of 3.2 cm. By comparing the results with the 25 cm measurements, Pippard confirmed that the asymptotic value of the surface resistance varies with frequency as ω^2 for a given material; this law is found to be accurately obeyed. At the same time Pippard was able to estimate the surface reactance X by an indirect method involving the superconducting behaviour of the metal. He found that $X_\infty/R_\infty \approx 1.76$ for tin, in quite as good agreement with the theoretical value $\sqrt{3}$ as may be expected. These experiments are considered further in § 4.72.

4.7. *Extensions of the theory.* We conclude by considering briefly various generalizations of the basic theory given above.

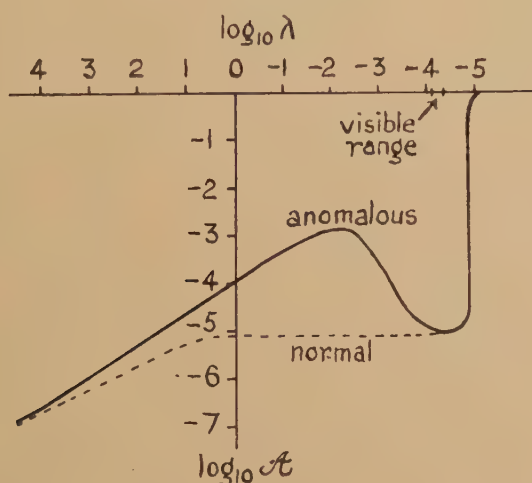
4.71. *Relaxation effects.* Up to now it has been assumed that the product $\omega\tau$ is small compared with unity. In the 9400 Mc/s experiments this condition is no longer satisfied, but we have seen that the results are in good agreement with the non-relaxation theory. This is explained by the extension of the theory to take account of relaxation, which may be obtained by replacing the time of relaxation in the Boltzmann equation (66) by a complex time of relaxation $\tau_i(1+i\omega\tau)$. It is then found (Reuter and Sondheimer 1948, § 6) that the integrals involve, instead of $i\alpha$, the generalized parameter $\xi = i\alpha(1+i\omega\tau)^3$: so long as $|\xi| \gg 1$, however, the final expressions obtained for Z are identical with (82) and (83), so that relaxation effects are negligible. The criterion $|\xi| \gg 1$ for neglecting relaxation is equivalent to the condition, which may be easily understood physically, that the period $1/\omega$ of the applied field shall be large compared with, not τ , but the time taken by an electron to traverse the skin depth, which is very much smaller than τ . As a result it is not to be expected that relaxation effects will play any part at wavelengths greater than about $\frac{1}{10}$ mm.

So long as the displacement current is negligible, the absorption coefficient \mathcal{A} of a metal for normally incident radiation, defined as the ratio of the intensities of the absorbed and incident waves, is directly proportional to the surface resistivity. According to the usual relaxation theory \mathcal{A} varies as $\sqrt{\omega}$ at low frequencies (this is the Hagen-Rubens relation), but is independent of frequency when $\omega\tau \gg 1$. The difference in relaxation behaviour which the anomalous skin effect introduces at low temperatures is illustrated in fig. 12, which shows the theoretical

absorption coefficient of a good silver specimen at liquid-helium temperatures and at frequencies ranging from long radio waves to ultra-violet light. The anomalous effects are most marked in the long infra-red, while at higher frequencies the relaxation effects tend to restore the normal behaviour; the maximum value of \mathcal{A} in fig. 12 is 1.5×10^{-3} and occurs at a wavelength of 70μ .

The absorptivity of pure electropolished metals for infra-red radiation of wavelength 14μ has recently been measured by Ramanathan (unpublished) at liquid-helium temperatures. The observed behaviour is not in agreement with the theory: the absorptivity in all cases exceeds the theoretical value by a factor of the order of 10, and is in fact about as great as the value obtained by extrapolating the $\omega^{2/3}$ law which holds when relaxation effects are negligible. The origin of the discrepancy

Fig. 12.



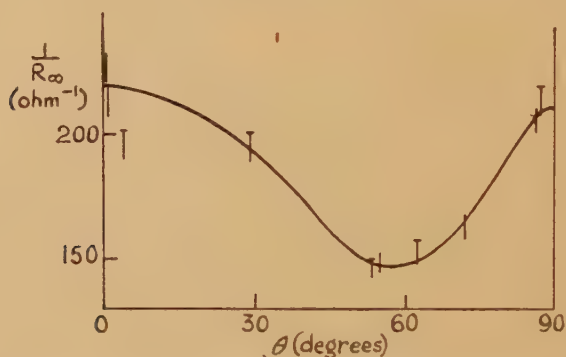
Theoretical absorption coefficient of silver at liquid-helium temperatures for $p=1$, plotted logarithmically against the wavelength λ .

is not clear at present, but it is possible that the disagreement with the relaxation theory is only apparent and is connected with the value of the surface reflection coefficient p . p was assumed to be unity by Reuter and Sondheimer when evaluating their formulae for high frequencies; it has been reported that in the infra-red region the absorptivity for $p=0$ is much higher than for $p=1$ (Holstein, unpublished calculations), but details of the computations are not yet available.

4.72. *The anomalous skin effect in anisotropic metals.* In his experiments at 9400 Mc/s, Pippard (1950 a) studied the surface resistance of single crystals of tin as a function of orientation. In fig. 13 $1/R$, measured under extreme anomalous conditions, is displayed as a function of θ , the angle between the tetrad axis of the crystal and the axis of the (cylindrical) specimen, which is also the direction of current flow.

The form of the anisotropy of $1/R$ is interesting, and differs from the result to be expected for a metal obeying Ohm's law. In an anisotropic metal the usual form of Ohm's law, $\mathbf{J} = \sigma \mathbf{E}$, must be generalized by treating σ as a symmetric second-order tensor. The theory of the normal skin effect in such a metal has been considered by Pippard, who has shown that, for the conditions of his experiments, it leads to a monotonic variation of R with θ , which is at variance with the experimental results. Pippard has pointed out that the reason for the anomalous behaviour is to be found in the form of the equations connecting \mathbf{J} and \mathbf{E} in the normal and anomalous theories. In the former the relation is a linear point-relation whether the metal is isotropic or not, the value of \mathbf{J} at any point being determined solely by the value of \mathbf{E} at that point. In the anomalous theory, on the other hand, the relation between \mathbf{J} and \mathbf{E} , while still linear, is no longer a point relation: there is then no longer any quantity analogous to σ which may be simply generalized by the introduction of a tensor, and as a consequence the mathematical formulation of the problem

Fig. 13.



Variation with orientation of surface conductivity of single crystals of tin.

permits a much greater freedom of behaviour in an anisotropic metal than is allowed by a point-relation. It is clear, furthermore, that the free-electron model which we have always employed hitherto will be entirely inadequate to deal with the present problem, and the anisotropic behaviour will depend on the anisotropic form of the energy surfaces in real metals. A first discussion of the problem using the ineffectiveness concept has been given by Pippard, but further investigation, both experimental and theoretical, is required before the phenomenon can be said to be fully understood.

4.73. *Application to superconductivity.* Most of the recent investigations on the high-frequency behaviour of metals at low temperatures have been concerned primarily with the properties of the superconducting state, a topic which we have ignored hitherto and which is outside the scope of the present article. It should be mentioned, however, that the

methods used in the theory of the anomalous skin effect find an application in the theoretical analysis of certain models of a superconductor, the so-called two-fluid models. The essential feature of these models is that the electrons in a superconductor may be thought of as belonging to either of two groups, superconducting and normal respectively, and that the current density may be written as the sum of separate contributions from each group, the supercurrent and the normal current. Under the influence of an electric field the normal electrons in a superconductor are supposed to behave in the same way as the electrons in a normal metal, and they will therefore exhibit similar free-path phenomena at high frequencies. The theory has been worked out in detail by Maxwell, Marcus and Slater (1949) for the particular model of a superconductor which is due to H. London (1940), and a careful critical discussion of the whole subject has been given by Pippard (1950 b), whose paper should be consulted for further details. It must be emphasized that, since there is as yet no accepted and fully developed theory of superconductivity, the status of these calculations on superconductors is much more speculative than that of the theory which applies to normal metals.

4.74. *Magnetic effects.* We wish to point out, finally, that effects analogous to those considered in § 3 are to be expected at high frequencies if a steady magnetic field is superimposed on the specimen. In particular, if the magnetic field is perpendicular to the surface of the metal, the surface impedance will vary with H in a manner which may be analysed theoretically by combining the methods of § 3.4 with those of the present section; at the same time there will be a high-frequency Hall electric field, and at low temperatures there will be an 'anomalous Hall effect'. The analysis of these phenomena, though laborious, is straightforward in principle, but elaborate calculations are probably not worth undertaking until experimental data have become available.

My best thanks are due to Mr. R. G. Chambers for several instructive discussions and for reading and commenting on the article in manuscript.

REFERENCES

- ANDREW, E. R., 1949, *Proc. Phys. Soc. A*, **62**, 77.
 CHAMBERS, R. G., 1950 a, *Proc. Roy. Soc. A*, **202**, 378; 1950 b, *Nature, Lond.*, **165**, 239; 1951, *Ph.D. Dissertation*, Cambridge.
 DINGLE, R. B., 1950, *Proc. Roy. Soc. A*, **201**, 545.
 DRUDE, P., 1904, *Ann. Phys.*, **14**, 936.
 EUCKEN, A., and FÖRSTER, F., 1934, *Nach. Ges. Wiss. Göttingen*, **1**, 43, 129. See also papers by various authors in *Ann. Phys.*, 1937–41.
 FAIRBANK, W. M., 1949, *Phys. Rev.*, **76**, 1106.
 FUCHS, K., 1938, *Proc. Camb. Phil. Soc.*, **34**, 100.
 LONDON, H., 1940, *Proc. Roy. Soc. A*, **176**, 522.
 LOVELL, A. C. B., 1936, *Proc. Roy. Soc. A*, **157**, 311.
 MACDONALD, D. K. C., 1949, *Nature, Lond.*, **163**, 637; 1950, *Proc. Phys. Soc. A*, **63**, 290.
 MACDONALD, D. K. C., and SARGINSON, K., 1950, *Proc. Roy. Soc. A*, **203**, 223.
 MAXWELL, E., MARCUS, P. M., and SLATER, J. C., 1949, *Phys. Rev.*, **76**, 1332.

- MOTT, N. F., and JONES, H., 1936, *The theory of the properties of metals and alloys* (Oxford : University Press), p. 268.
- NORDHEIM, L., 1934, *Act. Sci. et Ind.*, No. 131 (Paris : Hermann).
- PEIERLS, R., 1934, *Z. Phys.*, **88**, 786 ; *Helv. Phys. Acta*, **7** (Supp.), 24.
- PIPPARD, A. B., 1947, *Proc. Roy. Soc. A*, **191**, 385 ; 1950 a. *Ibid.*, **203**, 98 ; 1950 b, *Ibid.*, **203**, 195.
- PIPPARD, A. B., REUTER, G. E. H., and SONDHEIMER, E. H., 1948, *Phys. Rev.*, **73**, 920.
- REUTER, G. E. H., and SONDHEIMER, E. H., 1948, *Proc. Roy. Soc. A*, **195**, 336.
- SONDHEIMER, E. H., 1950, *Phys. Rev.*, **80**, 401.
- THOMSON, J. J., 1901, *Proc. Camb. Phil. Soc.*, **11**, 120.
- TITCHMARSH, E. C., 1937, *Introduction to the theory of Fourier integrals* (Oxford : Clarendon Press), chapter II.
- WILSON, A. H., 1936, *The theory of metals* (Cambridge : University Press).

*On the Generation of Vacancies by Moving Dislocations**

By Prof. FREDERICK SEITZ,
University of Illinois, Urbana, Ill., U.S.A.†

ABSTRACT

New experiments of Molenaar and Aarts, Blewitt and others seem to confirm the view of the author, previously based only on the experiments of Gyulai and Hartly and Stepanow on sodium chloride, that vacant lattice sites, and possibly interstitial atoms, are generated during plastic flow in ductile crystals, particularly in metals. It is pointed out that the average temperatures near a moving dislocation are probably not sufficiently high to evaporate vacant lattice sites or interstitial atoms as a result of thermal effects alone. Instead, one apparently must conclude that the imperfections are generated either by purely geometrical means during the looping of dislocations about appropriate obstacles, as the result of dynamical instability in the motion of a dislocation, possibly near a jog, or in the very high thermal pulses or 'spikes' which are generated either in the zone where two dislocations of opposite sign annihilate one another or near impediments where dislocations are strongly curved. It is pointed out that a pair of vacancies is probably stable near room temperature and may diffuse more rapidly than a single vacancy. It is also proposed that vacancies retained during quenching of Al-Cu alloys and those generated by cold-work play an important role in the precipitation process. The origin of work hardening in single crystals is discussed and several alternative interpretations, which involve the impediment of Frank-Read generators either directly or indirectly as a consequence of the generation of vacancies, are presented. The importance of prismatic dislocations formed by condensation of vacancies is restated. The role that vacancies formed by cold-work may play in determining the stored energy and decrease in density and in affecting processes such as creep and the hardening of latent slip planes is also discussed. Finally a few experiments are proposed, typical of those which could prove decisive in isolating the influence of vacancies.

§ 1. INTRODUCTION

IN a recent paper, the writer⁽¹⁾ has pointed out that experiments of Gyulai and Hartly⁽²⁾ and Stepanow⁽³⁾ on the influence of plastic flow upon the electrical conductivity of sodium chloride seem to imply that vacant lattice sites are generated within the crystal when dislocations move, as

* This work has been aided by research contracts with the United States Office of Naval Research and the United States Atomic Energy Commission.

† Communicated by the Author.

during ordinary 'static' experiments in the range of stress where plastic strain occurs. More recent experiments on metals by Molenaar and Aarts⁽⁴⁾, working at Druyvesteyn's Laboratory at Delft, which will be described briefly below, appear to support the same conclusion. There is enough ambiguity in the interpretation of the existing experimental work that the writer's point of view cannot be regarded as proved. Nevertheless, the implications of the viewpoint appear to be sufficiently far reaching if it is sustained by further research that additional discussion appears justified.

The first part of the paper will deal with an analysis of the experiments. Following this some speculative comments on the implications of the viewpoint will be made, with ample emphasis on the subjunctive mood.

The writer is indebted to Dr. T. Blewitt of Oak Ridge National Laboratory for calling his attention to the work of Molenaar and Aarts. Blewitt had similar experiments in progress at the time the work of the Netherlands group appeared. He had observed the increase in resistivity of polycrystalline copper with cold-work at low temperatures, to be described in the next section.

§ 2. EXPERIMENTAL OBSERVATIONS ; THE DENSITY OF VACANCIES GENERATED

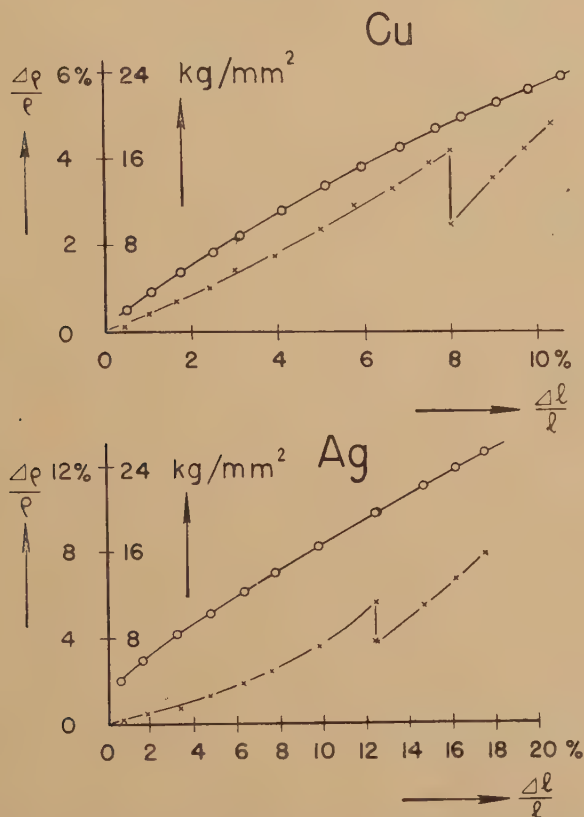
A. Experiments of Molenaar and Aarts

Molenaar and Aarts extended polycrystalline specimens of copper, silver and aluminium in tension by about 10% of strain at liquid air temperature and measured the change in electrical resistivity which accompanies the strain (see fig. 1). The specimens were then warmed to room temperature for various periods of time and cooled to liquid air temperature again. Measurements of electrical resistivity were continued along with further plastic flow. The principal qualitative results were: (a) The increase in electrical resistivity produced during the initial flow is at least partly eliminated by warming to room temperature. That is, the resistivity is lower when the specimen has regained liquid air temperature than it was after plastic flow, just prior to being warmed. Most of this annealing is achieved in ten minutes at room temperature. Practically all the resistance imparted to aluminium by the cold-work is annealed at room temperature, whereas only a fraction disappears in copper and silver. (b) The stress-strain curve does not appear to be altered by warming to room temperature. That is, the curve obtained when plastic flow is continued at low temperatures after the specimen has been annealed at room temperature, joins smoothly onto that obtained previously. Thus the curve showing electrical resistivity versus strain at liquid air temperature shows a discontinuity at the point where the strain was interrupted and the specimen was annealed at room temperature, whereas the stress-strain curve does not.

Blewitt (see § 1) has found that the recovery of electrical conductivity is apparent even at -80°C in copper.

From the viewpoint the writer presented previously⁽¹⁾, these results may be interpreted by stating that vacant lattice sites, and possibly other lattice defects, such as interstitial atoms, are generated during plastic flow and increase the electrical resistivity because they introduce centres which scatter the conduction electrons. We shall usually refer to the imperfections as vacancies for brevity, since the experiments on NaCl imply that a substantial fraction are of this type. More detailed

Fig. 1.



The stress-strain curves and increment in resistivity versus strain for copper and silver (after Molenaar and Aarts). The specimens were deformed at liquid air temperature and all resistivity measurements were made there. The discontinuities in the resistivity curves are the result of an 'anneal' at room temperature. The anneal is complete in the case of aluminium, which is not shown. Blewitt and investigators at North American Aviation Company have found that the annealing process takes place at a measurable rate in copper at -80°C .

work may show that interstitial atoms are as important. The vacancies are immobile at liquid air temperature, but have enough mobility at room temperature, and in fact even at dry-ice temperature, that they form aggregates which are sufficiently large to produce less scattering than the

isolated vacancies formed at low temperatures. This aggregation proceeds farther in aluminium than in copper or silver, since vacancies are more mobile in aluminium. The clusters may be in the form of planar sheets, probably nucleated at imperfections, which are the equivalent of dislocation rings. The linear dimensions of the 'rings' may be small initially.

If this interpretation is accepted as a tentative working hypothesis, it is possible to estimate the density of vacancies which are generated, at least to order of magnitude. Attention will be focused on copper, although the conclusions drawn seem to be equally valid for silver and aluminium.

When the specimen of copper is loaded to 24 kg per mm² (2.4×10^9 ergs per cm³) the strain achieved is 10%. Since the stress-strain curve is nearly linear, the work done is about 1.2×10^8 ergs per cm³.

During the strain, the resistivity is altered by about 6% (fig. 1), or rather would have been if the specimen were not annealed at room temperature. Since the resistivity of copper at liquid air temperatures is 0.322×10^{-6} ohm-cm, the change $\Delta\rho$ in resistivity is about 0.019×10^{-6} ohm-cm. Now Linde⁽⁵⁾ has determined the influence of various alloying elements upon the resistivity of copper and has found that the change varies linearly with the amount of a substitutional alloying agent added, as long as the amount of the alloying agent is not more than a few per cent. Moreover, the change in resistivity per atom per cent of solute varies with the nature of solute, being smaller for elements which resemble copper and larger for those which lie farther away in the given row of the periodic chart. The range of $\Delta\rho$ extends from zero to 10×10^{-6} ohm-cm per atom per cent of addition depending on the solute. We shall assume that a vacancy (or an interstitial atom) produces a relatively radical change in the lattice and assume that one atom per cent of vacancies would change the resistivity by 10×10^{-6} ohm-cm. It follows from this assumption that 10% of strain induces about 0.0019 atom per cent of vacancies, or about 1.6×10^{18} vacancies per c.c.

Since the energy expended in straining the lattice is 1.2×10^8 ergs per cm³, it follows that one vacancy is produced for each 7.5×10^{-11} ergs, or for each 47 eV. If we employ the rough rule⁽⁶⁾ that half of the energy of activation for self-diffusion in metals is associated with generation of vacancies and half with migration, we conclude that about 1.1 eV is required to generate a vacant lattice site in copper under conditions of thermal equilibrium, since the best value for the activation energy for diffusions is 48 kg cal per mol (2.1 eV per atom). Hence about 2% of the energy of cold-work is expended in producing vacant lattice sites, if the picture presented here is consistent. Taylor and Quinney⁽⁷⁾ have found that between 5 and 10% of the energy of cold-work is stored in copper during cold-work. The present work suggests that an appreciable fraction of this, that is a fifth or more, resides in the form of vacant lattice sites or interstitial atoms. In fact, it is possible that the majority of the stored energy resides in the form of vacancies. This

may be particularly true when single crystals are deformed in a manner approaching pure shear, for which asterism is very small.

There is another profitable way of viewing the foregoing results. A strain of 10% can be generated by having one dislocation move across every tenth plane. Since there are about 5×10^7 parallel planes per cm^3 in copper, it follows that in moving unit distance a dislocation produces about 5.3×10^{11} vacancies per unit length. Or it produces one vacancy in each plane of a set parallel to the slip vector in moving 7000 atom distances.

B. Experiments of Gyulai and Hartly

Gyulai and Hartly⁽²⁾ observed that the electrolytic conductivity of NaCl was raised by a factor of 100 or more when a specimen was loaded in compression by 2.5 kg per mm^2 (2.5×10^8 ergs per cm^3). The specimen was strained by almost exactly 10%, so that the energy expended was about 2.5×10^7 ergs per c.c.

If the writer assumes that the natural specimens of rock salt employed by these investigators contained one part per million of divalent impurity, which is a common value⁽⁸⁾ for 'pure' rock salt, the increase in conductivity implies that the plastic flow raised the density of positive ion vacancies to about 1 per 10^4 sodium ions, or raised the density to about 2×10^{18} per c.c. It is possible⁽⁹⁾ that the specimen was ten times purer, in which case the density of vacancies would be nearer 2×10^{17} . In the first case the energy expended per vacancy would be 6 eV, whereas it would be 60 eV in the second case. Wagner and Hantelmann⁽¹⁰⁾ have found that the equilibrium value of the energy of formation of a pair of positive- and negative-ion vacancies is 2.1 eV. Evidently the second value of 60 eV per vacancy during plastic flow is more reasonable than the first, for otherwise we would conclude that most of the energy of cold-work is stored.

In any case, it would seem to be highly desirable to have the experiment of Gyulai and Hartly duplicated under conditions in which the results are specified more completely, so that this relatively large ambiguity may be removed.

C. The Experiments of Masima and Sachs

Masima and Sachs⁽¹¹⁾ have measured the change of electrical resistivity of single crystals of brass (Cu70Zn30) with plastic flow and have correlated the changes with those in hardness and density. The relationships were also investigated during annealing stages at various temperatures. Although these experiments are exactly of the type that is needed to extend the topic under discussion in this paper and provide a test of the speculations, it is unfortunate that the metal chosen is the alloy brass, for single crystals of brass exhibit a number of irregularities, such as the Bauschinger effect, which are not observed generally in crystals. There is a strong probability that the effects observed in brass are somewhat atypical.

The outstanding results of the measurements seem to be the following:

(1) The electrical resistivity of the single crystals increases during plastic flow, the increase being of the order of 1% of the room temperature value during the first few per cent of plastic flow. The crystals were strained at room temperature, so that the results cannot readily be compared with those of Molenaar and Aarts, who deformed the materials at lower temperatures. The increment in resistivity was found to be independent of the temperature of measurement at any temperature where annealing did not occur. Since the resistivity of brass at room temperature is about 50 times larger than that of copper at liquid air temperature, it follows that the change in resistivity of brass during the first 10% of strain is about 50% of the resistivity of copper at liquid air temperature. In other words, the increment in resistivity is several times larger than the change observed in polycrystalline copper for comparable deformation.

(2) The increment in electrical resistivity produced in brass by cold-work does not rise with the stress-strain curve, but appears to saturate after about 10% of plastic flow. However, the resistivity rises again when the specimens, which were deformed in tension, begin to neck. The stress-strain curve rises during the entire process of plastic flow.

(3) The increase in electrical resistivity could be removed by annealing at a lower temperature than is required to resoften the specimens. For example, most of the increase in resistivity could be removed by annealing at 200° C for one half hour, whereas an anneal near 500° C was required to produce resoftening in a comparable period of time.

(4) The measurements of the change of density of deformed single crystals show that a decrease of the order of 0.06% can be obtained by a strain near unity. This would correspond to a generation of about 5×10^{19} vacancies per cm^3 , if we were to ascribe all of the change to the generation of vacant sites. Actually most of the change in density occurs only after double slip has begun to take place within the specimens: it is less than 0.006% during the first 10% of strain.

(5) The decrease in density starts to anneal with the same ease as the increase in electrical resistivity. Thus the two processes appear to be closely correlated. Actually the manner in which the density anneals is somewhat peculiar. The change in density of the cold-worked specimen with annealing temperature, for a half-hour anneal, rises as this temperature increases to about 200° C. It then falls in the range between 200° C and 400° C and rises again at higher temperatures. The investigators associate this second rise, which begins at about 400° C, with the onset of recrystallization. However, the original density does not appear to be regained even if the specimen is heated as high as 700° C, where dezincification begins. In other words, it does not seem likely that the annealing process consists merely in the coagulation of vacant lattice sites into larger and larger aggregates in the range below 500° before resoftening and recrystallization occurs.

Although the effects observed in these measurements are closely associated with those described in the preceding paragraphs of this section, it is evident that there probably are important side effects which may arise from the fact that brass is an alloy or that it is a highly anisotropic material. The alloy character may have a strong influence upon the events which occur during plastic flow, for the degree of short-range order may decrease, as the relatively large change in resistivity with cold-work implies is the case. In addition it is possible that vacant lattice sites are entrained in the lattice by inhomogeneities associated with fluctuations in composition.

The observation that the density decreases relatively rapidly when double slip begins suggests, at least vaguely, that the dislocations produce vacancies more readily when they encounter major obstacles than when they move through the nearly perfect lattice; however, it seems unwise to push this conclusion very far.

It would be desirable to have the experiments of Masima and Sachs repeated on single crystals of pure metals, particularly if the strain was carried out at low temperatures. The decrease in the amount of recovery of density which occurs as the annealing temperature is increased from 200° C to 400° C is a highly interesting effect. It would be valuable to know if it is a general phenomenon.

D. *Experiments of Tammann and Co-workers*

Tammann and Dreyer⁽¹²⁾ have carried out a series of somewhat unrefined measurements on the annealing of the increase in electrical resistivity and hardness induced by cold-work somewhat similar to those of Masima and Sachs. They subjected copper, silver, gold, palladium and platinum to a series of temperature anneals at room temperature and above. The specimens had been cold rolled or drawn by amounts varying from 10 to 98%. In the case of copper, silver and gold it was found that the increase in hardness and resistivity of a given specimen decreased in nearly the same manner as the annealing temperature was raised, whereas the annealing of hardness required a relatively higher temperature in the case of palladium and platinum. Results similar to that for palladium and platinum were obtained in iron and nickel by Tammann and Moritz⁽¹²⁾. The initial increase in resistance was in the same range in all cases.

Apparently the effect observed by Molenaar and Aarts and Blewitt could not be observed by Tammann and Dreyer in copper, silver and gold because the temperature at which their experiments started was sufficiently high that they had complete recovery from the effects the writer has associated with the production and coagulation of dispersed vacancies before measurements were started. Presumably their experiments on these metals are concerned primarily with the annealing of the effects arising from dislocations. The latter may be the dislocations responsible for plastic flow and for the generation of vacancies, or may be dislocations formed by precipitation of the vacancies generated during plastic flow,

which may act as impediments to motion of the original dislocations (see later sections).

On the other hand, the experiments on palladium, platinum, iron and nickel suggest that vacancies do not coagulate at room temperature in these metals and that one must heat well above room temperature in order to obtain the type of condensation observed in copper, silver and aluminium near room temperature or below. As we shall see later, it is possible that impurities present in these metals prevent the vacancies from coagulating by acting as traps which retain them in widely dispersed form.

Tammann, Dreyer and Caglioti⁽¹²⁾ have also investigated the influence of annealing on the recovery of electrical resistivity and hardness of alloys. The fractional increase in resistance may either increase or decrease with rising concentration of solute. However, the temperature required to produce a given amount of annealing always increases with rising concentration of solute. It seems to be a general rule that the increase in electrical resistance of the alloys may be annealed more easily than the increase in hardness.

The experiments on the effect of cold-work on the alloys may be more difficult to interpret than those on the relatively pure metals for reasons discussed in paragraph C. A part of the recovery of resistance and hardness may be related to the re-establishment of local order and hence be related only indirectly to the influence of vacant lattice sites.

§ 3. THE MODE OF GENERATION OF VACANT SITES

There seem to be four methods by which vacant lattice sites could conceivably be generated during plastic flow :

(1) By local heating of the lattice in the immediate vicinity of the dislocation as it moves through the lattice and absorbs energy from the applied stress field.

(2) By purely geometrical means in which dislocations of opposite sign moving in neighbouring planes annihilate one another and produce a row of vacancies.

(3) As a result of instability of the in-phase motion of atoms during passage of a dislocation. For example, it is possible that incipient vacancies are occasionally torn loose from a jog in a dislocation because of a local disturbance in the lattice.

(4) As a result of large, transient thermal pulses in regions where dislocations annihilate one another and produce a large lattice disturbance. We shall consider these processes individually.

A. Local Heating

It seems very unlikely that local heating over the entire length of the moving dislocation can be the agent effective in producing vacancies by evaporation, for the average temperature is probably very low. In order to proceed with a calculation of this temperature, we shall make several assumptions. First, we shall assume that the moving dislocation

achieves velocities near that of sound⁽¹³⁾ in the medium, but somewhat lower. Leibfried⁽¹⁴⁾ and the writer⁽¹⁵⁾ pointed out independently that a moving dislocation should dissipate the energy which it receives from the stress field at an appreciable rate. It appears to be very difficult to evaluate this loss precisely, but seems safe to say that the velocity of a dislocation becomes no larger than a value of the order of 10% of sonic velocity. Leibfried has focused attention on the dissipation which arises from the interaction of the moving dislocation with thermal waves, whereas the writer has considered the influence of anharmonic forces, which will cause dissipation even in an ideal, quiescent lattice, that is, even at a 'classical' absolute zero of temperature. Actually the argument of interest to us at the moment is not critically dependent upon the question of whether or not the dislocation actually may achieve sonic velocity, but this matter will be important in later discussions of this section.

The rate at which a unit length of a dislocation receives energy from the stress field is given by

$$s = \sigma a v, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where σ is the component of the stress field effective in inducing motion, a is the slip distance, which for simplicity we shall take to be the interatomic distance, as if the lattice were simple cubic, and v is the velocity of the dislocation⁽¹⁶⁾. If all this energy is transformed into thermal energy at the immediate vicinity of the dislocation line, the dislocation will act like a moving line source of thermal energy. We shall estimate the temperature distribution under steady-state conditions by the simple expedient of treating the dislocation as if it were a stationary line source and as if the medium had a simple thermal conductivity κ , whose value will be discussed below. The thermal field is then symmetrically distributed about the dislocation line. This simplified procedure probably leads to an underestimate of the temperature near the dislocation for a given assumed value of κ , since the thermal agitation will have a tendency to accumulate in the wake of the moving dislocation because its speed approximates that of sound. We shall attempt to compensate for this effect, which would be very difficult to treat rigorously, by underestimating the value of κ .

Under these assumptions, the temperature T at a distance r from the dislocation line (r is the radial variable in cylindrical coordinates in which the axis of the cylinder is coincident with the dislocation line) is

$$T = (s/2\pi\kappa) \log (R/r) + T_a \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

in which s is the rate of dissipation of energy, given by (1), and R is the distance at which the temperature may be regarded as having the ambient value T_a . R may conceivably be as large as the linear dimensions of the specimen, but its value is not critical for this discussion.

Let us consider a case in which σ has a value of 1×10^9 dynes/cm², as during the initial deformation in copper, a is 2×10^{-8} cm and v is 10^5 cm/sec (the velocity of compressional waves in a metal such as copper

is 3.9×10^5 cm per sec). Hence s is 2×10^6 ergs per cm-sec. If κ is chosen to be as small as 10^{-3} cal/(cm²-sec)(deg/cm) or

$$4.2 \times 10^4 \text{ ergs/(cm}^2\text{-sec)(deg/cm),}$$

the temperature rise in the vicinity of the dislocation attains a value

$$\Delta T = 8 \log (R/a)(^\circ\text{C}) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

in which r has been replaced by the interatomic spacing. Since $\log (R/a)$ should be less than 20, ΔT is less than 200°C . Actually we may anticipate much lower values if the assumption that the temperature field has rotational symmetry about the axis of the dislocation line is not grossly in error, for we have given all of the parameters values which would provide an upper limit for ΔT . For example, κ should be nearer unity than 10^{-3} cal/(cm²-sec)(deg/cm) in a good metal, and is actually larger than this at low temperatures. The value is 1.3 in conventional units in copper at 80°K . Even if use of the lower value of κ is justified in the immediate vicinity of the dislocation, where there is a high degree of lattice distortion, it would not be justified at somewhat larger distance, of the order of 10^{-7} cm, so that the value of 20 selected for $\log (R/r)$ is artificially large, by a factor of 5 or more.

The larger value of κ is more appropriate for salts, at least in the vicinity of room temperature. For example, κ is 0.021 cal/(cm²-sec)(deg/cm) at room temperature for NaCl. However, ΔT has been grossly overestimated even in this case. To consider one point, the stresses required to produce 10% of strain in NaCl under the conditions employed by Gyulai and Hartly are almost exactly ten times smaller than those employed by Molenaar and Aarts for copper, which served as the basis for the preceding estimate. Hence a more appropriate value of s would be ten times smaller in the case of NaCl, that is nearer 2×10^5 ergs/cm-sec.

We may expect vacancies to evaporate from a given atomic site on the edge of a dislocation at a rate

$$v = v_0 \exp (-\epsilon/kT) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

as a result of temperature alone. Here ϵ is the energy required to form a vacancy under equilibrium conditions and v_0 is of the order of 10^{13} sec⁻¹. If we set $\epsilon = 1.1$ eV, for the cases of NaCl and Cu, and assume that T is 500°K , that is, about 200° above room temperature as a result of local heating, the rate of evaporation is about 10^2 sec⁻¹. Since a dislocation moves unit distance in about 10^{-5} sec under the conditions we have assumed, the rate of evaporation, arising purely from temperature, is found to be negligibly small, that is, the dislocation would have to migrate of the order of 10 metres to evaporate one vacancy in a given plane of atoms normal to the axis of the dislocation, instead of a distance of the order of 10^{-4} cm. Since this calculation is based on the highest value of ΔT which we have estimated, namely 200°K , it seems very unlikely that the local heating of the dislocation is the important effect either in sodium chloride or in copper.

H. Brooks (private communication) has pointed out to the writer that temperatures much higher than the average may be generated at special places along the dislocation line, where it is held up temporarily by an impediment and then moves on as a result of the concentration of applied stress through curving of the dislocation in the direction of motion (Mott and Nabarro, Bristol Conference⁽¹⁶⁾). Two factors may contribute to the enhancement: First, after the stuck dislocation has been placed in motion, it may move for a distance in a stress field, associated with the impediment, which is much higher than the applied field; second, the curved dislocation may shorten its length after it has been freed from the restraint and thereby release some of the energy stored during lengthening. Both of these processes will be localized in relatively specialized regions; however, these regions may vary during the course of plastic flow. For example, the impediments may be dislocations associated with a system not parallel to the family to which the dislocation under consideration belongs and which change positions during plastic flow.

Of the two factors contributing to the enhancement of temperature, that arising from shortening of the dislocation line is presumably the most interesting, for the region in which the shortening takes place could be outside the region in which the impediment occurs, that is, where the crystal is relatively perfect. Actually we shall see that the temperatures are significant only within ten atomic distances or so of the impediment.

The conditions under which appreciably higher temperatures may be generated in this way can be envisioned from the following argument. If a dislocation moves one atomic distance a in simple cubic lattice, the energy transferred to it from an applied stress field of intensity σ is σa^3 per atomic spacing along the length of the dislocation. On the other hand, if a dislocation line which is bent into a half-circle of radius R straightens itself to length $2R$, the energy released is $(\pi-2)R\epsilon_0$ in which ϵ_0 is the energy per unit length of the dislocation line. If this energy is spread uniformly over the area $\pi R^2/2$ the energy transmitted per atom is $2(\pi-2)\epsilon_0 a/\pi(R/a) = 0.73\epsilon_0 a/(R/a)$. When σ is 10^9 ergs/cm³, σa^3 is about 0.01 eV in a typical case. Since $\epsilon_0 a$ is presumably at least 1 eV in a typical solid, the energy transmitted per atom in the two cases is about equal when $R/a = 100$. Thus R/a must be of the order of 10 if the energy released per atom in this way is to be an order of magnitude larger than that released by a stress field of 10^9 dynes/cm². The same result may be obtained in the following way. The stress field required to maintain a segment of a dislocation line curved into an arc with radius of curvature R is $\mu a/R$, where μ is the shear modulus. Since μ is of the order of 10^{11} dynes/cm², the effective stress field causing the dislocation to straighten is 10^9 dynes/cm² when R/a is about 100.

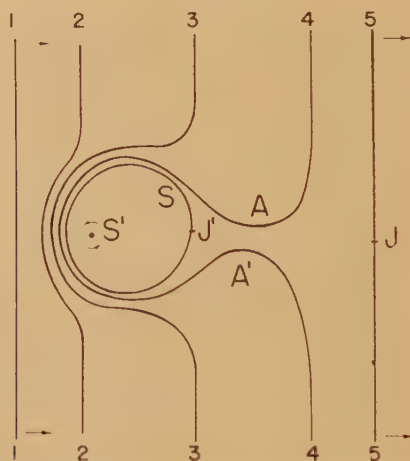
To summarize, temperatures of the order of 1000° K or higher can be expected only if R/a is small compared with 10. Thus the effect described by Brooks becomes important principally when the dislocation is held

up by impediments which produce a very high local curvature, that is, by impediments in which the stress opposing the motion of the dislocation is at least 10^{10} dynes/cm². The high temperatures are then realized only within ten atom distances or so of the impediment. The most important instance of this might be that in which impurity atoms are distributed throughout the lattice and cause large local curvature of dislocations.

B. Geometrical Means

It is evident that there are many purely geometrical methods by which dislocations can generate vacancies or interstitial atoms. One such method is sufficient for consideration. Let us assume that a screw, or Burgers, dislocation meets another screw dislocation which extends in a direction oblique to the plane in which the first screw dislocation is moving (fig. 2). We may assume that the slip vector for the second

Fig. 2.



Generation of vacancies or interstitial atoms by interaction of two dislocations (schematic). S' represents the intersection of a stationary screw dislocation with the slip plane in which another dislocation is moving. The line of the second, moving dislocation is indicated at successive positions during motion by 1, 2, 3, 4, 5. Initially the moving line lies entirely in a single slip plane. However, after looping around the stationary screw dislocation and leaving a closed segment behind, it possesses a jog J , on either side of which it lies in different slip planes. The ring left behind will also have a jog, J' . The segments A and A' , which are shown just before combining, move in neighbouring slip planes and have the character of Taylor dislocations if the moving dislocation is initially of screw type (stage I). Hence a row of vacancies or interstitial atoms will be generated when A and A' meet, depending upon the sign of the dislocation involved in the collision.

dislocation is not in the same direction as that of the moving direction, so that it can be regarded as stationary. The moving dislocation will be restrained by the stationary dislocation in the region where the latter

intersects the slip plane of the former and will, as a result, loop around it and eventually pinch off a segment which encircles the stationary dislocation. During the process of looping, the portions of the moving dislocation which are extended parallel to the direction of motion become, in effect, lengths of edge, or Taylor-Orowan, dislocations. Portions of the loop on opposite sides of the stationary screw dislocation will be edge dislocations of opposite sign, so that they will attract one another. However, they will be moving in neighbouring atomic planes because of the screw character given to the slip planes by the stationary screw dislocation. Hence when the edge-type segments of the moving dislocation meet they will annihilate during the process of pinching-off, but will leave either a row of vacant lattice sites or a row of interstitial atoms, depending upon the relative orientation of the Burgers vectors associated with the two dislocations. If we postulate that a row of vacant sites produced in this way will break up into component vacancies and disperse somewhat as a result of the local temperature engendered during the annihilation of the edge dislocation (see § 4 below), we obtain a feasible, essentially geometrical method of generating vacant sites.

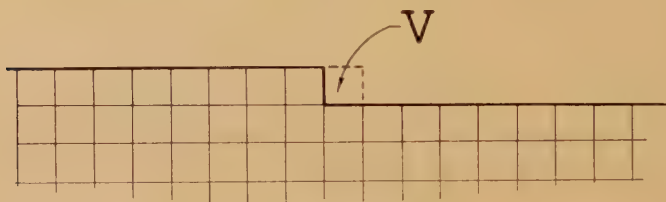
Since the density of dislocations which may be generated in this way, or in similar, purely geometrical ways, depends very strongly upon the entire pattern of dislocations present in the lattice, it is difficult to estimate the density which can be achieved. Speaking very roughly, we might expect a given dislocation to produce 10^4 vacant lattice sites or interstitial atoms in moving a distance of the order of the linear dimensions of mosaic blocks, namely about 10^4 atomic distances. For we might expect the line of the moving dislocation to undergo major distortion in moving from one mosaic region to another and hence to produce an aggregate extra length of the order of 10^4 atom distances, at least part of which will be annihilated in the manner described above. Hence a given dislocation would generate one vacancy per atomic plane normal to the axis of the dislocation in moving about 10^4 atomic distances. This number corresponds very closely to the observed value, as estimated in § 2. However, it is difficult to place much confidence in the estimate since we are relatively ignorant about the general distribution of dislocations.

C. Dynamical Instability

The passage of a dislocation through a crystalline lattice at a speed approaching that of sound requires detailed coordination of the positions of the atoms of the lattice under conditions in which inertial forces are high. As the dislocation approaches a given atom, the atom is accelerated, attains a maximum velocity when it is at the centre of the dislocation, and then comes to rest again by a more or less symmetrical process of deceleration. The local strain near the centre of the dislocation is of the order of 0.1, so that the forces between atoms are no longer harmonic. It is conceivable that dynamical instabilities occur along the dislocation line at the high velocities which probably occur and that vacant lattice

sites are produced as a result of this instability. This might occur with particular ease at a point where a dislocation jogs from one slip plane to a parallel one which is separated by one atomic distance. Such jogs are the seat of what the writer has termed⁽¹⁾ incipient vacant lattice sites. These sites are particularly easy to visualize when one is dealing with Taylor dislocations, so that the dislocation can be regarded as formed by the introduction of an extra plane into the lattice in a direction normal to the slip vector. The jog then corresponds to a position where the extra plane has a step and the incipient vacancy occupies the vacant site at the step (fig. 3). It is geometrically possible for a step of this type to move in the direction of the Burgers, or slip, vector along with the remainder of the dislocation. However, it is possible that instabilities develop and that the atom which is to move into the position of the incipient vacancy occasionally fails to do so. The vacancy would thereby be left behind. Since the energy required to free the incipient vacancy and make a normal vacancy of it is large, of the order of 1.1 eV in copper or sodium chloride, it follows that this instability would require the

Fig. 3.



Lateral view of the 'extra plane' in a Taylor dislocation and position of incipient vacancy V. The heavy line represents the Taylor dislocation when viewed in a plane normal to the slip plane and normal to the slip direction. The squares represent atomic cells. The Taylor dislocation possesses a jog and the incipient vacancy C occurs at the jog (dashed square).

cooperative action of a number of atoms in the neighbourhood of the jog, which would share in some communal manner the energy required. One might expect vacancies to be produced preferentially, if this mechanism prevails, in any crystal in which the energy required to generate a vacancy is appreciably less than that required to generate an interstitial atom.

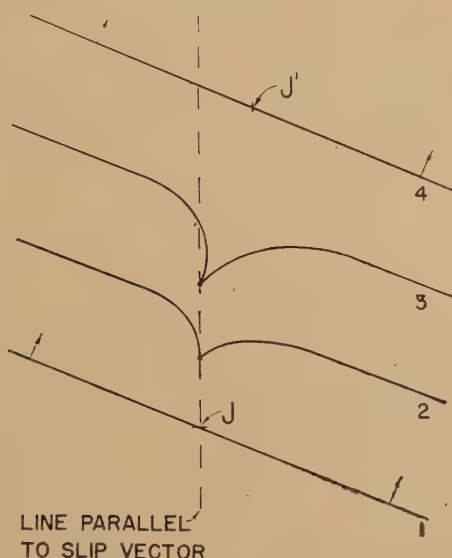
It is also possible⁽¹⁷⁾ that the instability arises from the fact that the jog becomes caught in the lattice and is brought to rest, or nearly to rest. The adjacent part of the dislocation may then loop around the restrained area and, as in the case of a dislocation that has looped around a static screw dislocation discussed in § 2 above, generate a row of vacant sites or interstitial atoms (fig. 4). In the present case the jog furnishes the differences in position of the slip planes on which the annihilating segments of the dislocation move, so that the screw character of the stationary Burgers dislocation is not required. Evidently this method of generating

vacancies passes into a purely geometrical method when the dislocation is stopped by a rigid impediment in the region near the jog. It differs from the case discussed in § 2 (fig. 2), only in the means by which the displacement of two parts of the dislocation to neighbouring slip planes occurs.

D. Influence of Transient Thermal Pulses

A large amount of stored energy is released whenever two dislocations moving in the same or closely neighbouring, parallel slip planes unite by annihilation. Since the energy per atomic length of a dislocation is of the order of 1 eV or more, the energy released is of this magnitude per

Fig. 4.



Generation of vacancies or interstitial atoms from a jog in a dislocation by geometrical means. The dislocation of mixed character possesses a jog J initially (line 1). As it moves in the direction indicated by the arrows, the jog becomes caught (lines 2 and 3), and the line loops around the impediment. After the loops meet, the jog is at position J' and is displaced relative to the rows of atoms in the direction of the Burgers vector (dashed line). A row of lattice vacancies or interstitial atoms, equivalent to the distance between J and J' normal to the dashed line, is generated in this process. The sign of the imperfections generated depends upon the sign of the jog.

atomic plane normal to the dislocation line. Hence the line at which the dislocations combine becomes the seat of a very large source of thermal energy which raises the local temperature to the neighbourhood of $10,000^\circ \text{K}$.

The thermal surges pointed out by Brooks (paragraph A), which occur when a dislocation line having sufficiently large curvature is straightened,

the energy of annihilation per atomic length along a pair of dislocations is comparable to this, it is probable that the peak temperature attained is somewhat lower, when expressed in units of eV, for the annihilation energy will be liberated over the distance, of the order of five atomic spacings, for which the strain in a dislocation is large. Hence the Boltzmann factor in the expression giving the rate of generation of vacancy-interstitial pairs should be somewhat less than unity, perhaps of the order of 10^{-2} . In addition, the members of a pair will exert a very strong attractive force on one another because they produce lattice strains which are both large and of opposite sign. The energy of interaction falls off as the third power⁽²⁰⁾ of the distance for sufficiently large distances of separation and probably maintains this mathematical form to distances of the order of a few atomic spacings. Hence pairs which are separated by a few atomic distances, which is as far as they can be separated during the time the thermal pulse endures, will probably migrate together and recombine. This is certain to be the case if the ambient temperature is sufficiently high to allow the vacancy or interstitial atom some degree of mobility as is true near room temperature in sodium chloride and copper, for jumps which bring the two closer together should be much more probable than those which separate them, as long as they are separated by no more than a few atomic distances.

To summarize, the thermal bursts accompanying the annihilation of dislocations should make it possible for rows of defects of a given kind that are generated as a result of the geometry of annihilation to disperse by a few atomic distances. A few pairs of vacancy-interstitial atoms may be formed in the thermal burst; however, it probably is necessary to maintain the crystal at very low temperatures to prevent them from recombining as a consequence of normal thermal diffusion in the subsequent period of time because of the strong attractive forces.

§ 4. THE MOBILITY OF VACANCIES AND PAIRS

A. *Alkali Halides*

The study of ionic conductivity in the alkali halides has contributed a substantial amount of information concerning the ability of these units to migrate. It is known⁽²¹⁾, for example, that the positive-ion vacancy is much more mobile than the negative-ion vacancy. The ratio of the jump frequencies⁽²²⁾ is about 10^5 at room temperature in KCl, and the value is probably close to this in NaCl. The activation energy for migration of the positive-ion vacancies is⁽²³⁾ about 0.80 eV in NaCl and is⁽²⁴⁾ about 0.85 eV in KCl. Hence the jump frequency is found to be near 1 sec^{-1} near room temperature, if the values obtained at higher temperatures are extrapolated to lower temperature. This value is supported by the observation that processes involving ionic migration actually occur at room temperature.

Essentially nothing is known about the behaviour of interstitial ions in the alkali halides. On the other hand, interstitial silver ions are⁽²⁵⁾

very mobile in the silver halides. It is probable that interstitial alkali metal ions would possess a mobility comparable to that of the positive-ion vacancies; however, the energy required to generate interstitial ions in the alkali halides is larger than that required to generate vacancies.

Vacancies of opposite sign should have an association energy of the order of 1 eV in the alkali halides. In fact, calculations by Reitz and Gammel⁽²⁶⁾ have led to a value of 0.89 eV in the case of NaCl. It seems very likely that the coupled pair of vacancies⁽²²⁾ possesses a relatively high mobility, perhaps higher than that of the positive-ion vacancy. Dienes⁽²⁷⁾ has estimated that the activation energy for migration of this pair is near 0.4 eV in KCl, in contrast with a value nearly twice this for the positive-ion vacancy.

B. Copper

The precise value of the activation energy for migration of vacancies in copper is not known. However, it is probably close to 1.0 eV, since the entire energy required⁽²⁸⁾ to form a vacancy and permit it to diffuse is only 2.1 eV and it seems reasonable, in light of the calculations of Huntington and the writer, to take the viewpoint that the activation energy for migration is about half this total. On the assumption that the jump frequency is determined by an expression of the form (4), it is readily found that this frequency is 10^{-4} at room temperature and 10^{-21} at 100° K. Thus isolated vacancies should possess some degree of mobility at room temperature, but be frozen into position at liquid air temperature.

It is interesting to speculate on the question of the energy of association of vacancies in metals. There is little doubt that vacancies have a short-range attractive force, since any degree of clustering should diminish the ratio of surface to volume of the void formed by the vacancy. For example, the average kinetic energy associated with the wave functions of the conduction electrons should diminish if the electrons are compelled to avoid only the cavity associated with a coupled pair of vacancies instead of the two voids associated with two isolated vacancies. It is possible, however, that the vacancies repel one another at large distances because they produce identical elastic distortion in the lattice. It would be highly interesting to know the general form of the interaction potential, but until this question has been studied in some detail we can only guess the results.

It is perhaps not unreasonable to suppose that the energy of association of a pair of vacancies is close to 1.0 eV, so that nearly half the energy of about 2.0 eV associated with the pair of vacancies when they are separated is regained by combination. If this is the case, the total activation energy required to separate a pair should be near 2.0 eV, since the total activation energy would be roughly the sum of the activation energy required for migration of an isolated vacancy plus the energy of association. The time required for the pair to dissociate under thermal fluctuations at room temperature would then be infinite for all practical purpose. A direct

theoretical value, obtained with the use of an equation of the form of (4) is 10^{20} sec.

There is no reason to suppose that the energy of association of a pair of vacancies actually is quite as large, or larger, than the energy of formation of a single vacancy. However, if this situation actually did occur, the energy required to form a pair of vacancies would be equal to or less than the energy required to form a single vacancy. In this case, pairs would probably be formed in preference to single vacancies by normal thermal means, since the entropy increase associated with the presence of a pair is probably larger than the entropy increase associated with a single vacancy. As Brooks has pointed out to the writer (private communication), a crystal in which pairs are formed in preference to single vacancies could exhibit remarkable properties. For example, if diffusion takes place by means of vacancies, the activation energy for self diffusion might be markedly lower in such a crystal than in a comparable one in which single vacancies are preferred, for, as we shall see below, it is not unreasonable to suppose that pairs of vacancies diffuse more rapidly than single vacancies.

As in the case of the salts⁽²⁷⁾, we should expect the pair to be very mobile, for the number of repulsive bonds, arising from closed-shell interaction that must be overcome is reduced. If we assume that the activation energy for migration of the pair is near 0.5 eV in copper, the jump frequency at room temperature is about 10^5 sec^{-1} , whereas that at 200° K is about 1 sec^{-1} . The jump frequency at 100° K is of the order of $10^{-10}/\text{sec}$, which is negligibly small. It is evident that pairs of this type may play a very important role in coagulation affected by migration of vacant lattice sites at room temperature, and even at 200° K . In fact, it is natural to suggest, at this very incomplete stage of our knowledge, that the diminution of resistivity observed in the experiments of Molenaar and Aarts, when the specimens deformed at liquid air temperature are annealed at room temperature, is the result of coagulation of vacancies by diffusion of pairs, or possibly larger aggregates which also possess a relatively high mobility.

Although the calculations of Huntington and the writer indicate that the energy required to form an interstitial atom in copper is much larger than that required to form a vacancy, the results also suggest that the activation energy for diffusion of the interstitial atom by replacement of neighbours, that is by the interstitialcy⁽⁶⁾ mechanism, is of the same magnitude as that for migration of vacancies. Hence interstitial atoms should have some mobility at room temperature. However, even if pairs of interstitial atoms become bonded, we should not expect such pairs to diffuse faster than single interstitial atoms. For the situation is presumably just the opposite of that around a pair of vacancies in the sense that the pair of interstitial atoms will compress the immediate neighbourhood in which they are situated and make migration more difficult. It seems unlikely that the migration of interstitial atoms plays

an important role in copper and similar metals in the vicinity of room temperature.

Diffusion in aluminium will be considered in the next section.

§ 5. INFLUENCE OF COLD-WORK ON THE RATE OF PRECIPITATION⁽²⁹⁾

A. Experiments

It appears to be well established⁽³⁰⁾ that cold-work increases the rate at which copper precipitates in aluminium containing a few per cent of copper. This effect has been investigated most carefully by Gayler⁽³⁰⁾, whose observations may be summarized as follows.

(1) The measurements show that the logarithm of the time required to achieve a given state of precipitation, such as the time t_m required to achieve the second, or maximum, peak in hardness, varies linearly with $1/T$, where T is the absolute temperature. Gayler has found that this line is displaced parallel to itself by cold-work produced during rolling. The investigations were made for reduction in area lying between zero and 90% and for temperatures between 100° C and 187° C. This result suggests that the activation energy for the process determining precipitation is not altered by cold-work. Instead, the rate process is enhanced by a factor which is essentially independent of temperature. If we assume that diffusion is the limiting process, the result supports the view that the coefficient of the Boltzmann factor in the conventional expression for the diffusion constant, and not the activation energy, is altered by cold-work.

(2) The rate of ageing at a given temperature is a continuous function of cold-work. Gayler has shown that the curve giving the relation between the reduction of cross section and the logarithm of time required to achieve the secondary maximum of hardening is approximately linear for reduction to about 80%. The curve then flattens near 100% reduction when plotted in this manner. The result might have been plotted in other ways which would have been more revealing. The writer has attempted to transcribe Gayler's results in such a way that $1/t_m$ is represented as a function of the strain $\epsilon = -\log(1-R)$, in which R is the fractional reduction in cross section. The resulting curve is nearly linear for the entire range of R . Thus, within the precision that this transcription of Gayler's data allows, one may say that the increase in rate of precipitation varies linearly with the strain to strains of the order of 3.2 (96% reduction).

B. Vacancy Diffusion

Gayler's results may be explained on the assumption that diffusion of the copper atoms responsible for precipitation occurs with the use of vacancies in the manner proposed by Johnson and Wagner⁽³¹⁾ and that the vacancies produced during cold-working are able to assist in the process. The precipitation rate is determined by the density of vacancies. From the slope of the $\log t_m$ versus $1/T$ curves given by Gayler, which is

available for three decades of t_m , one may estimate that the activation energy for diffusion of copper is about 28,000 cal per mol. This is smaller than the experimental values⁽³²⁾ of the diffusion coefficient of copper determined by various investigators, which range between 33,900 and 41,900. These values are determined from diffusion data obtained at higher temperatures than those of interest in normal precipitation studies; moreover, the range of temperature employed is sufficiently small that the diffusion coefficient varies only by a factor of 10, which does not permit very accurate evaluation of the activation energy. We should expect the activation energy for precipitation after cold-work to be lower than that for diffusion at higher temperature, for the vacancies will presumably be at thermal equilibrium at elevated temperature and the energy required to form a vacancy near a copper atom will appear in the activation energy for diffusion, whereas this is not the case if vacancies are formed during deformation. It is very interesting to note, however, that the activation energy for precipitation in the range below 200° C in the absence of cold-work is the same as that after cold-work. From the present point of view, this is to be explained on the assumption that the density of vacancies is higher than the equilibrium value during precipitation. There would appear to be two possible explanations of this fact:

(1) Some of the vacancies formed at elevated temperatures are captured by copper atoms during the quenching step and succeed in coming more nearly to complete equilibrium only after the precipitation process has taken place. Before being bound into dislocations or other stable clusters they succeed in helping the copper atoms to precipitate. An argument to be given in a later paragraph of this section indicates that the density of vacancies retained in Gayler's undeformed specimens is of the order of 3×10^{16} per cm^3 , that is, of the order of one per million aluminium atoms. This is the density which would be at equilibrium at about 500° C, that is, at a temperature somewhat above those employed in Gayler's experiments. If this interpretation is valid, the activation energy involved in precipitation should increase at temperatures somewhat higher than those employed in Gayler's experiments because the density of vacancies will correspond to the equilibrium value. Moreover, the temperature at which the equilibrium value is attained should increase with increased cold-work because the density of vacancies arising from cold-work increases.

(2) The vacancies which are responsible for precipitation in the undeformed alloys are generated during precipitation because of localized plastic flow which accompanies the process. Presumably the stresses engendered during precipitation will be relieved, at least in part, by the motion of dislocations in regions about the precipitates. One might expect vacancies to be produced during this process and enter into the diffusion process until they are recaptured by a dislocation, or cluster together. This explanation can be valid only if the density of vacancies formed during precipitation is independent of the rate of precipitation. For if the alloys in which precipitation is rapid have a higher density of

vacancies than those in which it is slow, one would not expect the rate of precipitation to be proportional to the value for a cold-worked specimen, which presumably starts with a fixed density of vacancies, over the entire range of temperature.

On the whole, the first of these two possibilities appears to be more reasonable than the second, although available experimental material is not sufficient to distinguish between the two.

C. Nabarro's Viewpoint

Nabarro⁽³³⁾ has suggested that the increase in rate of precipitation with cold-work, without a change in activation energy for the process, is related to an increase in the density of nucleating centres for the precipitation process. It would be very interesting to test the relative merits of this proposal with that given above by comparing the size of the precipitate particles in a deformed and an undeformed specimen, for this size should vary inversely as the density of nuclei, and hence be strongly dependent on cold-work, if Nabarro's proposal is correct.

The specimens employed by Gayler contained 4% of copper. Since the density of vacancies in cold-worked specimens should be less than this even when the reduction in area is 96%, corresponding to a strain of about 3.2, it is evident that the vacancies must act in the manner of catalysts if we are correct in assuming that the Johnson-Wagner mechanism prevails. Presumably the vacancies deposit a copper atom at the site of a precipitate and then wander away to assist another somewhat in the manner in which positive-ion vacancies presumably catalyse the coagulation of F-centres in coloured alkali halides⁽²²⁾. Again, it is quite possible that the vacancies migrate as pairs or larger clusters during the process. The activation energy for self-diffusion in aluminium is apparently unknown, but is probably in the vicinity of 35,000 cal per mol. If we assume that the activation energy for diffusion of a single vacancy is about half of this, the jump frequency should be about 10^3 at 100°C. This rough value is sufficiently large that we cannot exclude the possibility that the vacancies act as single units instead of pairs or clusters. Moreover, it is sufficiently large to suggest that the vacancies generated in aluminium during cold-work are able to migrate more than a few atomic distances at temperatures near room temperature, in agreement with the observation of Molenaar and Aarts that the increase in electrical resistivity induced in aluminium during cold-work at low temperatures anneals completely at room temperature in a time of the order of a few minutes.

D. Density of Retained Vacancies

There is one additional matter of interest that may be gleaned from Gayler's measurements. The rate of precipitation of copper is about 1000 times faster in the specimen which has been reduced by 96% than in the material which has been subject to quenching without cold-work. If we assume that the first specimen has been strained by 3.2 and that the density of vacancies is proportional to the strain, we may conclude

that the cold-worked specimen contains about 3×10^{19} vacancies per cm^3 . It follows that the quenched specimen, which was not cold-worked must have contained about 3×10^{16} vacancies per cm^3 , or somewhat less than one vacancy per million normal sites. This would not be sufficiently large to determine by measurements of density and lattice spacing, even if the complication of distortion as a result of the formation of precipitates did not exist.

E. Rohner's Theory

In surveying the literature on precipitation in the aluminium-copper system, the writer noted the paper of Rohner⁽³⁴⁾ with interest. This investigator has suggested that vacancies play an important role in the precipitation and hardening process, however, his views do not appear to follow the conventional pattern of current thought. He proposes, in effect, that the copper atoms migrate interstitially and that at least a part of the hardness of the alloy is a result of the formation of vacant sites as a consequence of the transfer of copper atoms from normal to interstitial positions. The evidence presented in the foregoing paragraphs suggests that the vacancies produced in the way envisaged by Rohner would coagulate in a few minutes at the temperatures where precipitation is observed, so that they should not have an influence on hardening. In justice to Rohner, one must admit that something akin to Frenkel disorder may occur in aluminium and that interstitial diffusion may be an important process. Our present knowledge of diffusion in the aluminium system does not preclude this possibility.

§6. SPECULATIONS ON THE PHENOMENON OF WORK-HARDENING

A. Geometry of Slip

The observation, discussed in §2, that a substantial part of the energy stored during plastic flow may be associated with the vacant sites or interstitial atoms generated during motion of dislocations raises the question of whether the lattice defects formed in this way could have an important effect on the work-hardening of solids. We shall discuss this topic in a highly speculative manner, as our present state of ignorance requires.

The experiments of Heidenreich and Shockley⁽³⁵⁾ and of Brown⁽³⁶⁾ on aluminium suggest that plastic flow in single crystals takes place in the following way:

(1) Flow begins by the generation of groups of dislocation rings on a few atomic planes of the lattice. The generators are sections of dislocations which run in the given slip plane in a more or less perfect manner for a distance between 10^3 and 10^4 atomic spacings and then move out of the plane. The terminal points in the plane act as anchors. As Frank and Read⁽³⁷⁾ have demonstrated, such segments may act as the generators for many dislocation rings. The planes on which generation starts may be widely separated from one another during the early phases of plastic flow.

(2) A given generator spins out a sufficient number of dislocations that a total slip distance of about 2000 Å, or about 1000 atomic spacings, can be realized. At least this many of the total progeny of the generator arrive at the boundary of the crystal and produce a step. Additional dislocations may be formed by a given generator, but they presumably do not get to the surface of the crystal. The steps correspond to the slip lines observed under high magnification on specimens which have been deformed plastically. To avoid confusion with other terminology, we shall refer to the steps as slip lines in this manuscript even though the shear motion takes place over a plane.

(3) As plastic flow continues, dislocations are generated in slip planes parallel to but displaced from those in which it first started, thereby producing parallel slip lines. Moreover, there is a tendency for a number of such planes to group together, forming equally spaced lamellae which constitute a *slip band* or *slip zone*. The total displacement in each subsequent slip plane seems to be about the same as in the first, namely 1000 atomic distances. There is some indication⁽³⁸⁾, however, that this distance actually decreases in successive slip lines. The spacing between bands and the number of slip planes or lines in a given band varies with the temperature⁽³⁹⁾. At least this is the result obtained by Brown from an analysis of grains in polycrystalline aluminium. The most common interplanar spacing of lines in a band is 200 Å, a distance first observed by Heidenreich and Shockley, but larger values were found at higher temperatures. Both the spacing between bands and the number of slip lines in a band decreases with decreasing temperature. For example, the spacing between bands is about 2μ at room temperature and there are about four slip lines per band. The spacing decreases to less than half this at liquid air temperature and there are only one or two slip planes per band.

(4) X-ray measurements by Kellar, Hirsh and Thorp⁽⁴⁰⁾ with a very fine beam, obtained by use of a capillary slit system, confirm the fact that rolled aluminium contains 'crystallites' of the order of several microns in size. These presumably are the regions between slip bands. On the other hand, the measurements also indicate that the crystallites are somewhat distorted. The distortion decreases with a reduction in particle size, presumably as a result of polygonization⁽⁴¹⁾, if the specimens are maintained at room temperature for many weeks. This result suggests that some dislocations occur in the region between slip bands and aggregate with long annealing. There does not seem to be reliable information available on the fraction of plastic flow associated with the dislocations between slip bands, but we shall assume for definiteness that it is not an important part. Further work may show that this assumption is erroneous.

It should be emphasized that the most reliable and detailed observations⁽⁴²⁾ of the slip bands on the scale of the electron microscope in published form are restricted to aluminium, although similar results have

been obtained on brass by Craig and Czyzewski⁽³⁸⁾. A relatively drastic assumption is involved in extending the results found for aluminium and brass to other metals. However, we shall proceed on this basis. Should subsequent observations show that the slip bands which occur in other solids are radically different in nature or that a large component of flow takes place by the action of single dislocations more or less randomly distributed instead of those generated in highly localized regions, many of the conclusions drawn below will need revision.

B. Origin of Work-Hardening ; Spiral Prismatic Dislocations

Work-hardening apparently originates in the fact that larger and larger stresses are required to place new generators of dislocations into operation. Hence the problem of explaining hardening would appear to raise two component problems: (a) Why do the generators lose their ability to produce new rings? (b) Why are stonger and stronger forces needed to set new generators into operation?

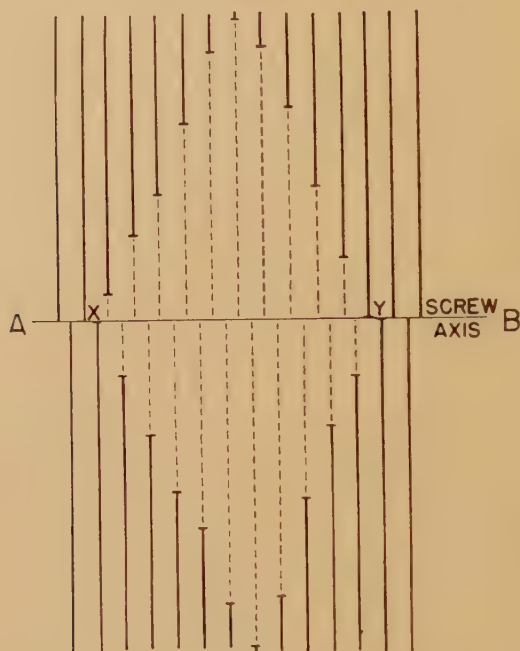
We shall restrict attention here principally to the problem of determining the factors which cause a given generator to lose its ability to generate new rings. However, a few comments on the second aspect of the problem seem appropriate.

The writer⁽⁴³⁾ has proposed elsewhere that the Frank-Read generators which are responsible for the slip lines in a given slip band may be associated with prismatic dislocations, formed by condensation of vacancies in sheets on a plane within the lattice. Moreover the equally spaced generators which give equally spaced slip planes in a given slip band or zone are associated with precipitated sheets of vacancies that wind around a screw dislocation and form a spiral prismatic dislocation (figs. 5 and 6). The quantitative aspects of such prismatic dislocations will be discussed in more detail in a later paper. They appear to have the required regularity if it is assumed that consecutive arms of the spiral are separated by 100 atom distances, which is consistent with a low value for the total energy of distortion associated with the spiral. Parallel segments of each turn of the spiral may act as generators. As is indicated schematically in fig. 6, it is reasonable to assume that the 'cone' on which the spiral winds has a polygonal cross section when segmented in planes normal to its axis, which extends in the direction of the screw dislocation on which the vacancies have condensed. A polygonal rather than a circular or elliptical cross section may be expected for the same reason that the stable habit of a crystal is polygonal instead of spherical or ellipsoidal, namely the energy of the configuration⁽⁴⁴⁾ is lower. Hence we shall suppose that the sections of each turn of the spiral which act as genertors are straight lines parallel to one side of the polygonal cross section. It follows that the length of the generating lines will increase linearly with the distance of the arm from the axis of the cone.

H. Brooks has pointed out to the writer (as yet unpublished work) that if the equilibrium form of prismatic dislocations are circular and

if the elastic properties of the lattice were isotropic, prismatic dislocations probably would not act as Frank-Read generators when in a shearing stress field. Instead, they would elongate initially in the direction of the applied stress, maintaining a constant projection on a plane normal to the Burgers vector. When this elongation has proceeded sufficiently far that the dislocation ring lies almost in the slip plane, segments of the

Fig. 5.



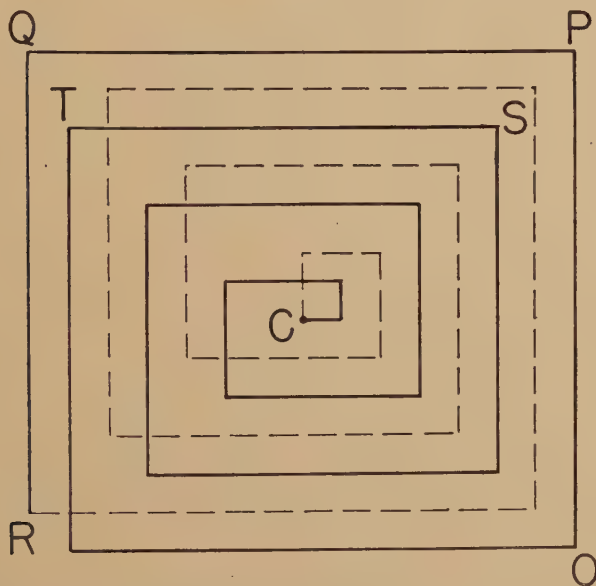
Schematic view of vacancy pattern in a spiral prismatic dislocation, formed by precipitation of vacancies. The horizontal line AB represents the line of a screw dislocation for which the Burgers vector is in the direction AB. The vertical lines represent the intersection, with the plane of the paper, of atomic planes which are transformed into a spiral sheet as a result of the presence of the screw dislocation AB. The lines jog at AB because the screw is centred on this line. The full segments of the vertical lines are sections of the planes which are complete, whereas the dashed segments are portions of the plane on which vacancies have precipitated and "eaten" a hole. In three dimensions, the dotted portion is enclosed in a double 'cone' formed by abutting the bases of two cones. The short horizontal bars outline the intersection this double cone makes with the plane of the paper. The common axis of the cones is the line AB. In practice the planes will collapse in the direction AB in order to eliminate the void represented by the dashed segments. Successive planes will be displaced by a relative distance equal to the Burgers vector for new dislocations. This is assumed to be an allowed translational vector of the lattice and need not be the same as the Burgers vector for the screw dislocation. The terminal points between the full and dashed lines (designated by horizontal bars) then become the loci of a spiral dislocation which winds about the double cone, starting at X and ending at Y.

ring parallel to the Burgers vector, which have screw character, will move in a direction normal to the Burgers vector. It is the writer's opinion, at the time this is written, that the anisotropic nature of crystal lattices will usually influence the equilibrium form of prismatic dislocations sufficiently that their behaviour will be as described in the preceding paragraphs.

C. Frank's Criticism

Two objections to the spiral prismatic dislocation have been raised by F. C. Frank (private communication):

Fig. 6.



Schematic view of spiral prismatic dislocation formed from precipitate of vacancies in fig. 5 after lattice collapses on precipitate to form dislocations. The diagram corresponds to the state of affairs when viewed along line AB. The intersection of line AB with plane of paper is point C. The double spiral, half of which is represented by full line and half by dotted line, represents the spiral prismatic dislocation corresponding to the locus of the thermal points of the planes in fig. 5 (heavy bars). The full line winds on the right-hand cone and the dotted line on the left-hand cone of the double cone formed from the vertical dashed segments in fig. 5. It is assumed that the cones have a polygonal cross section (square in this schematic case for simplicity) because the system will have lowest energy when the segments of the spiral dislocation run in highly specified crystallographic directions. Vertical and horizontal segments of a given turn of the dislocation, such as the segment OP or PQ, may act as generators in the sense of Frank and Read. It should be emphasized that the full and dashed arms of the spiral terminate at different points in the direction normal to the plane of the paper, both of which project at C in this drawing. The dashed line terminates at the point X in fig. 5, whereas the full line terminates at Y.

(1) Frank believes that at the time a sheet of vacancies collapses to form a dislocation, at an early stage of formation of a prismatic dislocation, the Burgers vector will not be an allowed translation of the lattice so that the lattice will be in registry over an entire plane and the dislocation will be one form of the type he terms sessile⁽⁴⁵⁾. Since a sessile dislocation of this kind will possess a higher energy than one in which the lattice is in registry over the area where the sheet collapses, provided the area of the sheet is sufficiently large, the writer assumes that the lower energy state is achieved during the cooling of the specimen. The instability which impels this conversion is particularly large in the case of a spiral prismatic dislocation since the energy of an appreciable volume of the lattice is at stake. Moreover, the distorted region associated with the nucleating dislocation is present to catalyse the conversion. In this case, the problem of achieving the lowest state, in which the dislocation is not sessile, is somewhat analogous to that of obtaining a precipitate which has the equilibrium lattice form. The attainment of this form is aided, in the case of the prismatic dislocation, by the fact that the precipitation process occurs within a hundred degrees or so of the melting point where the atoms are highly mobile.

(2) Frank believes that the inner arms of the spiral will be inhibited from acting as generators by the fact that they must cut across the outer ones. The writer recognizes that the outer arms offer a potential impediment to the operation of the inner ones, but believes that this obstacle is by no means insurmountable. It is to be observed, in the first place, that the outer arm is not subject to such a restriction, as the example shown in fig. 6 demonstrates, so that it may multiply freely. Parallel to this, it is interesting to note, on the experimental side, that Brown⁽³⁶⁾ has found that the first slip plane in a band occurs much more easily than the later ones. The following is one of several factors which can effectively mitigate the potential impeding influence of the outer arms of the spiral. Let us assume that the spiral has the form shown in fig. 6, and that the upper and lower horizontal segments are those which can multiply under the given applied stress, whereas the side segments act as the anchors, at least initially. The arms of the spiral have relatively broad freedom of motion under action of a uniform shearing stress, as the writer has emphasized in connection with the behaviour of a simple prismatic dislocation consisting of a single loop (see fig. 15⁽¹⁵⁾ in particular). In the first place, an arm or a segment of an arm may slide on a cylinder parallel to the Burgers vector. The upper and lower segments of an arm will tend to slide in opposite directions under the action of a uniform shearing stress. If the outer arm, such as OPQR in fig. 6, slides relative to the inner one, such as ST in fig. 6, in this way, it will increase the distance between the vertical anchor segments for it and for the next arm of the spiral. This distance may become comparable to the length of the horizontal segments which act as multipliers. Moreover, if the outermost arm becomes sufficiently extended along the cylinder, in the

direction of the Burgers vector (see fig. 15⁽¹⁵⁾), the vertical anchor segments will become stretched to a point at which they can be regarded as segments of dislocations in places parallel to the slip planes connected by jogs. When this point has been reached, large portions of the outer spiral arm may depart from the region of the prismatic dislocation under the action of the applied stress, leaving debris which consists of vacancies or interstitial atoms. This stage may occur after the outer peripheral arm has completed its life as a generator. Once the outer arm has departed, the next arm is in a position to act as a generator without undue difficulty. The dislocation rings produced by the next generator will have to pass through the debris left by the outer arm. It is possible that this can occur only at a temperature sufficiently high that thermal fluctuations are available to overcome highly localized obstacles, as Brown's results seem to imply.

D. Critical Stress for Generator

The critical stress required to make a given generator produce dislocations is probably determined by two factors: (a) Its length, which determines the radius of curvature of the generating line when it loops about the generate new dislocation rings. (b) The stress fields determined by neighbouring dislocations. If the stress associated with (a) is greatest, we should expect the critical stress to vary inversely as the length, since the stress required to move a dislocation⁽⁴⁶⁾ is inversely proportional to the radius of curvature. In terms of the properties of spiral prismatic dislocations, this would imply that successive generators in a given spiral would be set into operation as the stress is increased. If the arms of the spiral are labelled by integers, 1, 2, 3, 4, etc., as one passes from the outer periphery of the spiral to the axis, there should be a linear relation between the integer index of the generator which operates in producing a given increment of strain and the total strain, provided the spiral prismatic dislocations furnish the most important source of generators for extensive slip, for example, for a net strain near unity. It is possible, of course, that other dislocations, which are not associated with spiral prismatic dislocations, and which move much more easily, are responsible for the first stages of slip in a well-annealed specimen.

One may estimate the restraining stresses which the dislocations in a spiral exert on one another with the use of equations derived for one-dimensional lattices of dislocation lines. Cottrell⁽⁴⁷⁾ has shown that the stress required to detach one member of a row of equally spaced Taylor dislocations from the row, in the case in which the row lies in a plane normal to the slip plane, is $1.1G\lambda/2\pi(1-\nu)b$. Here G is the shear modulus, λ is the slip distance, ν is Poisson's ratio and b is the distance between the dislocations. This is of magnitude $10^{-3}G$ when b is about 100 atomic spacings. The corresponding value is near 5×10^8 dynes/cm² in metallic copper and about 3×10^8 dynes/cm² in aluminium. If the writer's model of a spiral prismatic dislocation is qualitatively correct, these values should apply at least in range of magnitude for the initial

ordered form of the dislocation, since parallel segments in neighbouring arms, which act as generators, should resemble closely a linear lattice of Taylor dislocations for which b is $100a$. In consequence, we should not expect the generators associated with spiral prismatic dislocations for which b is this small to operate until the stress is well above the lowest values of the critical shearing stress observed in pure, well-annealed single crystals. For the latter stress may be as small as 10^6 dynes per cm^2 . As noted in the previous paragraph, other dislocations will presumably determine plastic flow until the stress is at least 10^8 dynes per cm^2 . These may still generate single slip lines if they have the characteristic form required for Frank-Read multiplication, as for example, if they have the form of simple prismatic dislocations.

E. *Experiments of Röhmer and Kochendörfer*

It should be observed at this point that Kochendörfer and Röhmer⁽⁴⁸⁾ have found evidence to show that the stress-strain curve of single crystals of aluminium is linear when the specimen is deformed in pure shear, in contrast with the parabolic relation obtained in the classical experiment of Taylor⁽⁴⁹⁾ and his associates. The linear relation could be explained with the use of spiral prismatic dislocation under the following simple assumptions: (a) The generators associated with spiral dislocations are the principal ones determining the strain in the range extending to unity, studied by Kochendörfer and Röhmer. (b) Each generator produces about 1000 dislocations which move to the edge of the specimen as observed by Heidenreich and Shockley and Brown. (c) The stress required to start generation is determined principally by the length of the generator, so that the stress required to set in operation generators lying successively nearer the axis of the spiral dislocation increases linearly with the integer index that specifies the position of the arm in which the generator lies relative to the outer periphery.

F. *Experiments of Averbach and Warren*

Averbach and Warren⁽⁵⁰⁾ have found that filings of brass exhibit a very high degree of distortion, all atoms being appreciably displaced relative to one another. Presumably the strain in filings of this type is higher than unity. In any case, the results imply that the slip lines eventually merge with one another as strain continues, or at least the region between lines becomes strained, and the deformation becomes more nearly homogeneous.

G. *Cessation of Generator; Destruction*

Why does the generation process cease after the generator has produced about 1000 rings? The simplest explanation to offer is that the generator becomes systematically destroyed after it has formed this number of rings. We saw previously, in connection with the problem of interference of arms of the spiral prismatic dislocation (paragraph F), that a segment of a prismatic dislocation may become elongated in the presence of a

uniform shearing stress ⁽¹⁵⁾. For example, a simple prismatic dislocation may be stretched in the direction of the Burgers vector, opposite ends moving in opposite directions on a cylinder, or prism, parallel to this vector. Similarly, an arm of a spiral prismatic dislocation may be stretched in this way. During this stretching the anchor segments are converted into dislocation-lengths lying in neighbouring slip planes which are connected by jogs from one slip plane to the next. When the distance between jogs becomes sufficiently great, the prismatic dislocation, or section of a spiral arm, need no longer be confined to the surface of the cylinder and may expand, with the production of rows of vacancies or interstitial atoms. These imperfections will be primarily interstitial atoms if the prismatic dislocations have been formed by the precipitation of vacancies in sheets; for the increase of the projected area of the dislocation ring on a plane normal to the Burgers vector, which measures the number of vacancies associated with it, must be compensated by the generation of interstitial atoms. Once this stage is reached the dislocation will presumably cease to act as a Frank-Read generator, but will behave more nearly like a simple dislocation ring which lies entirely in the slip plane, differing from the latter because of the jogs which act as sources of vacancies or interstitial atoms. It is possible that the generator disappears abruptly ⁽⁵¹⁾ and with considerable regularity as a result of a transition of this kind after producing about 1000 rings.

The simplest objection to this mechanism seems to be that it does not give an obvious explanation of the hardening of latent slip planes (§8), unless we assume that the debris left by the generator when it vanishes is effective in inhibiting generators which would produce slip in latent planes. Moreover, it does not explain the observation (§7) that metals may be resoftened to their original high ductility at relatively low temperatures, as if the generators could be reconstituted easily.

In this connection it is interesting to observe that Haase and Schmid ⁽⁵²⁾ have found that zinc crystals which have been strained by 50%, and have then been permitted to stand for one day, recover their original softness (see §7), even after the process has been repeated several times. This experiment suggests very strongly that the generators are not destroyed during cold-work, but are jammed by a process which prevents them from forming new dislocations without producing any essential geometrical change.

In any case, we shall investigate explanations which depend upon the effects produced by vacancies generated during the migration of plastic flow, and shall place principal emphasis on explanations of this type for the present.

H. *Trapping of Dislocations by Vacancies*

We saw in §2 that each dislocation which passes across a slip plane seems to generate about one vacant lattice site per plane normal to the line of the dislocation in moving about 7000 atom distances. Since the vacancies produced in copper cannot migrate very far from the slip plane at temperatures near or below room temperature (§4), the vacancies

produced by one dislocation should be present to interfere with the motion of the next. After 1000 dislocations have passed, the density of vacancies in the vicinity of the slip plane should be very high. If one projects the vacancies onto the slip plane, the fractional density of vacancies should be of the order of 10% of the density of atoms in the slip plane. The vacancies would presumably be distributed within a few atomic distances of the slip plane. Since they will distort the lattice very strongly in this region, it is not surprising that the dislocations should encounter difficulty moving across the slip plane after a thousand have passed. On the contrary, it seems strange at first sight that such a large number may move through the plane without becoming stuck much earlier. There seems to be two possible explanations of this:

(1) The vacancies are not produced at random over the slip plane but are produced in highly localized groups. One such group may have a highly restraining action upon the passage of dislocations through a particular portion of the slip plane, so that they must loop around it; however, the other portions of the plane remain relatively free. Motion becomes difficult when groups overlap everywhere on the slip plane.

(2) The dislocations depend upon their latent kinetic energy of motion to carry them through the rough regions of the lattice. Initially the portion of the slip plane near the generator remains smooth so that there is a region in which the dislocations may be accelerated to a velocity where inertia is sufficient to maintain motion in spite of serious obstacles. They are stopped when the region of the slip plane near the generator where acceleration occurs becomes sufficiently rough that they cannot gain kinetic energy, or when the stress field in the vicinity of the generator drops to zero because of the back-reaction of dislocations which it has engendered and which have become caught.

It is difficult to decide between these two possibilities, if indeed, they are the only ones. Eshelby⁽¹³⁾ has pointed out that a dislocation which succeeds in achieving a velocity near that of sound possesses a kinetic energy near its "rest energy", that is, near the energy of distortion, which is of the order of 1 eV per atomic length along the line. If the stress field is of the order of 10^8 dynes cm^2 , the dislocation will take an energy comparable to its rest energy from the field in travelling about 1000 atomic distances. Hence it could emerge from the region where it is generated with sufficient velocity to overcome obstacles which do not dissipate its energy too rapidly.

Once a dislocation has come to rest in the region where there is a high density of vacancies, it will probably become locked into position because the vacancies in the vicinity will migrate to it and form clusters in order to relieve the local stress field in the manner suggested by Koehler⁽⁵³⁾ in another connection. Thereafter a very high stress would be required to set it into motion again, just as dislocations which have attracted carbon atoms to them become locked, in accordance with Cottrell's⁽⁵⁴⁾ theory of the upper yield point in steel.

I. Fouling of Generator by Imperfections

One of the questions which arises at this point is the following: Is it possible that the generator becomes fouled because of imperfections which are produced during the process of generation, other than the dislocation rings responsible for slip? Suitable fouling of this kind could be ample to explain the cessation of generation and, hence, work-hardening, at least in cases in which only one parallel family of slip planes operate. One naturally thinks of vacant lattice sites or interstitial atoms that may be produced within the area in which the multiplication process occurs and which may prevent the generator from moving.

Consider first the time associated with each generation cycle. If we assume that the generator is several thousand atomic spacings in length and that the dislocation moves with maximum velocity of 10^5 cm/sec, the time required for the generator to loop about the anchor points once would be 10^{-9} sec or longer. This is sufficiently greater than the time, of the order of 2.5×10^{-11} sec (see equation (6)), required for the heated zone produced during the annihilation phase to cool to ambient temperature, that we should expect the path in front of the generating segment to be at ambient temperature continually as the dislocation loops about the anchor points. Hence we should not expect vacant lattice sites to evaporate from it simply because it passes through a heated zone associated with the region where annihilation occurs. On the other hand, pairs consisting of vacant sites and interstitial atoms may be produced in the annihilation region because of this high temperature. Many of the pairs may recombine, but a few may become separated as a result of diffusion during successive waves of temperature.

We should not expect the generating segment to become entangled with a stationary screw dislocation which cuts obliquely across the slip plane. A dislocation of this type might occur occasionally in the area in which the generator operates, but this would be an atypical case. Thus vacancies presumably are not produced by purely geometrical means involving a stationary screw dislocation near the generator.

We seem to reach the conclusion that the generator will encounter resistance to motion as a result of vacant sites or interstitial atoms produced in the generating area only if (a) the imperfections are formed as a result of dynamical instability during motion of the generating dislocation in regions where it attains velocities comparable to the velocity of sound, or (b) if the vacancies and interstitial atoms are produced as pairs in the heated annihilation zone and diffuse apart. We know far too little about the mechanism by which the imperfections are generated by moving dislocations to evaluate the likelihood of the first possibility. It seems entirely feasible, however, that the limited life of a generator is determined by the fact that debris accumulates in its area of operation, particularly in the annihilation region, where it attains maximum velocity. Once it has been caught in this region, it presumably will become locked into position as a result of the migration to it of some of the products

generated during motion. Until it becomes stuck, the dislocations it produces may be essentially identical.

Even if the generator does not become fouled because of debris produced within the generating area, it will cease to be a good generator if it becomes surrounded by progeny which are locked into position. For a point will eventually be reached at which any further dislocation rings produced will be repelled by those which are frozen into position in the same place. The latter will have the same character and sign as the new dislocations in regions where they approach, so that repulsive forces are guaranteed.

To summarize, it seems easy to provide several reasons why the flow started by a given generator should cease after the generator has produced a certain average number of dislocations, if the moving dislocation engenders vacancies at the rate suggested by the experiments described in §2. Moreover, it is not difficult to understand the observation that the given slip plane does not become operative again during the period in which other unused generators are available. It is also evident, however, that our picture of the process is exceedingly rudimentary.

J. Influence on Subsequent Generators

The vacancies produced in one slip plane should not have an appreciable influence in planes separated from this plane by 100 atomic distances if they do not diffuse far. For the stress field arising from a point imperfection dies off inversely as the third power of the distance and should have a very minute value at distance of the order of 200 Å. Hence the vacancies produced by the dislocations derived from one generator should not have an appreciable direct influence on neighbouring generators. On the other hand, the stress field of dislocations falls off much more slowly, in fact, as the inverse first power of the distance. It might appear that the dislocations which are trapped near a given generator would have an appreciable affect on a dislocation 100 atomic distances away. However, this source of inhibition on the new generator apparently does not provide a major obstacle for slip in the neighbouring plane for the following reason. We would expect the trapped dislocations to have a stronger repelling effect on the generator in their own plane than on that in the neighbouring slip plane because the shear stress produced by a dislocation is greatest in its own plane. Moreover, the field which they produce at their own generator should be no greater than the applied field which prevailed when this generator was active. Thus the trapped dislocations should not be able to restrain the new generator when the applied stress is increased. In other words, the trapped dislocations will produce a field which compensates for part but by no means all of the applied force and should not be able to prevent new slip in the neighbouring plane.

K. Stored Energy

The following point is worth noting. It is not difficult to believe that the total stored energy associated with the trapped dislocations produced by a given generator is less than that associated with the

vacancies and interstitial atoms the dislocations form if the projected density of the latter on the slip plane is as high as ten atomic per cent. Even if the trapped dislocations were spaced one hundred atomic distances apart over the entire slip plane, the energy associated with them would be equal to that of the vacancies only if the energy per atomic length of the dislocation were ten times greater than that associated with a vacancy. This spacing of trapped dislocations is probably as high a value as could conceivably occur if slip takes place through the action of generators. For the restraining stress such closely spaced dislocations would exert on the generator would then be comparable to the applied field in the range of plastic flow normally studied, that is would lie near $10^{-3}G$. In short, it seems entirely conceivable that vacancies and interstitial atoms account for the great majority of the stored energy in single crystals in which plastic flow occurs in the simple manner described above.

L. *Magnetization Studies of W. F. Brown and Kaufmann*⁽⁵⁵⁾

W. F. Brown has observed that the approach to saturation of the magnetization curve of heavily cold-worked polycrystalline material can be explained on the assumption that the heavily worked material possesses a linear imperfection. The density of these imperfections attains a value of 10^{12} per cm in the specimens studied. The result suggests that strongly cold-worked materials contain a high density of dislocations. This conclusion is not necessarily in conflict with the general picture presented in this section, which places major responsibility for work-hardening during the early stages of strain upon the vacancies generated. In the first place it is possible that the distortional effects arising from trapped dislocations become more important than those arising from vacancies or interstitial atoms in very heavily cold-worked material (see paragraph F of this section). In the second place, it is possible that the vacancy clusters, which trap dislocations and impede the dislocations, act as a linear source of distortion in the sense required by Brown's experiments if they aggregate into small dislocation rings.

M. *Mott's Viewpoint of Hardening*

In an as yet unpublished Guthrie lecture, which the writer received while the present paper was in the final stages of preparation, Mott has examined the grosser slip markings on single crystals of aluminium studied by Cahn, with the aid of the optical microscope, and has drawn a number of stimulating conclusions, which will be commented upon briefly here. First, he concludes that the rings issuing from Frank-Read generators ultimately become much elongated in the direction of the Burgers vector, that is, in the direction in which the dislocation possesses edge character. He proposed that the screw segments are inhibited by the fact that they must frequently generate vacancies or interstitial atoms in order to encircle screw dislocations which run normal to the slip plane, although other explanations are possible. Second, he proposes that the grosser irregularities of slip bands may be explained in terms of the somewhat chance

encounters and combination of groups of dislocations issuing from different sources, which do not move in identical slip planes and hence can combine only with the production of jogs.

If the writer takes Mott's analysis at its face value, which seems highly appropriate, he concludes that there are only about 10^8 active Frank-Read generators per cm^3 in the 20% range of strain investigated by Cahn. This is about 10^4 times smaller than the total number of dislocation rings to be expected in a metal having normal mosaic structure and about 10^2 times smaller than the number of spiral prismatic dislocations one might anticipate if all rings were associated with such dislocations. This conclusion suggests either that only a very small fraction of the rings are geometrically able to multiply, or that the stress required to start multiplication varies from ring to ring and that only a small fraction can operate in the range of stress employed by Cahn. The writer is inclined to accept the second of these explanations and, in keeping with the tenor of the present paper, associate work-hardening with fouling of generators which can operate at low stresses, so that successively higher stresses are required to maintain plastic flow.

Mott, however, proposes that work-hardening is associated with the impediment which dislocations issuing from different generators exert on one another when they collide. This view is in general keeping with Taylor's older theory. However, there is the significant difference that the characteristic distances in Mott's theory are of the order of 10^{-3} cm instead of 10^{-5} cm. It is evident that a final decision between the viewpoints of Mott and the writer will require additional experimental and theoretical work, particularly the former. It is not unlikely that both mechanisms may contribute to hardening, and in fact the experiments⁽⁴⁸⁾ of Röhmer and Kochendörfer suggest that the form of the stress-strain curve may be very sensitive to the uniformity of the conditions of strain, as one might expect if the mechanism of hardening proposed by Mott contributed a component to the work-hardening curve which varies with the conditions of strain, and on which is imposed a more primary irreducible mechanism of the type described in this section.

§7. RESOFTENING OR RECOVERY OF WORK-HARDENED METALS⁽⁵⁶⁾

According to the picture developed in the preceding sections, we should expect the influence of work-hardening to disappear if the crystals are raised to a temperature sufficiently high to dissolve the aggregates of vacancies, provided they join the edges of dislocations, form larger dislocation rings, or recombine with interstitial atoms that are present. The dislocations which exist in the lattice might then move freely and produce further disorder. Presumably the vacancies would become dispersed throughout the lattice during this procedure, instead of being restricted to the immediate vicinity of the slip planes. Hence it may prove possible to restrain them at impurity atoms which have an attachment for vacancies and thus detect their presence by studying the rate of diffusion of foreign

atoms during the annealing procedure. The experiments of Tammann and co-workers⁽¹²⁾ on the influence of alloying agents on the annealing of the increase in resistivity suggest that such restraining action may be common.

The estimates of diffusion rates presented in §5 imply that the vacancies produced in aluminium may be able to migrate sufficiently rapidly in the vicinity of room temperature that they will eventually become dispersed and produce annealing through extensive combination with one another or with dislocations. This result is in agreement with the observation that pure aluminium will resoften at room temperature. Zinc also resoftens⁽⁵²⁾ appreciably in the course of a day at room temperature and it would seem to follow that the activation energy for migration of the vacancies in this metal should be somewhat less than the value of about 18,000 cal per mol estimated for aluminium in §5. Actually Cottrell and Aytakin⁽⁵⁷⁾ have measured the rate of recovery of single crystals of zinc and have compared the results with measurements of steady-state creep, to be described in the next section. They conclude that the limiting rate process is the same in the two cases and is near 30,000 cal per mol, that is, more nearly like the activation energy for self-diffusion. On the other hand, Boas and Schmid have obtained a lower activation energy for creep in zinc, so that the matter is not in a well-defined state at present.

It is interesting to note that Tietz, Anderson and Dorn, and Cherian, Pietrokowsky and Dorn⁽⁵⁸⁾, as a result of studies of resoftening in polycrystalline aluminium, have come to the conclusion that two types of imperfections are involved in the process. They differentiate between these on the basis of the relative ease with which they may be annihilated during annealing. They have referred to the two types of recovery, associated with the annealing of each of the hypothetical imperfections, as *metarecovery* and *orthorecovery*. The first type takes place at lower temperatures than the former. The discussion of the preceding sections makes it reasonable to suggest that metarecovery is associated with coalescence of vacancies in such a way as to free at least a fraction of the dislocations or generators from the impediment which prevents them from moving, whereas orthorecovery is associated either with complete removal of the vacancy or with the restoration of the original pattern of dislocations.

Evidence for two similar types of recovery has been obtained by Slifkin and Kauzmann (as yet unpublished) from studies of the transient creep of zinc single crystals. After a strain of about 10^{-4} , a one minute rest at 35° C produces a marked softening, as shown by a greater creep rate after the rest than before. The effects of this recovery disappear rapidly upon continued straining. Rest periods of up to 100 minutes duration produce little further recovery beyond that found after one minute. If the crystal is allowed to rest for 40 hours, however, complete resoftening to the original state is observed. The rapid, limited recovery and the slower, complete process correspond to the metarecovery and orthorecovery, respectively.

§8. HARDENING OF LATENT SLIP PLANES

A. *Experiments*

Experiments⁽⁵⁹⁾ demonstrate that latent slip planes which cut across the actual slip planes harden at about the same rate as actual slip planes. The most recent investigation⁽⁴⁸⁾ of this effect, namely that of Röhms and Kochendörfer on aluminium deformed in pure shear, suggests that the hardening σ_e of latent slip planes is related to that of the actual slip planes by the simple equation

$$\sigma_e = \sigma \{ \cos(g_1 N_2) \cos(g_2 N_1) + \cos(g_1 g_2) \cos(N_1 N_2) \}, \quad . \quad . \quad . \quad (7)$$

where σ is the stress required to produce flow in the actual plane, g_1 and N_1 are the direction vector for slip in the real plane and the normal to the real slip plane. g_2 and N_2 are the same quantities for the latent system.

Although it is not difficult to explain the qualitative features of hardening of the latent slip planes on a theory of the type discussed in the previous sections, it is apparent that there is no easy way of explaining the simple relation (7). Let us consider several points for orientation.

Let us suppose that we are interested in the range of strain between zero and unity which is investigated most commonly. It is clear that if slip is localized entirely on the slip planes responsible for slip lines, so that 1000 dislocations move across each on the average, these planes will not be equally spaced at a distance of 200 atomic spacings throughout the specimen. For in this case the total strain would be of the order of 10 instead of unity or less. If, for a strain of the order of unity, the bands which occur contain 10 slip planes on the average, the average spacing between bands would be 10^{-4} cm. In fact these are in the range of values found by Brown for slip in aluminium. It follows that there are areas between slip bands which are approximately ten times broader than the slip bands.

B. *Possible Sources of Hardening*

Two possibilities would seem to be feasible: Either the slip generators for latent planes are distributed independently of those for the active planes, in which case some would occur in the regions where there are no slip lamellas, or the generators for all types of slip planes are closely correlated and either lie in the same regions of space, or are, for unexplained reasons, identical units. In the first case we must conclude that the hardening of latent planes does not arise from the fouling of the generators for the latent planes by the slip lamellas. Instead, it originates in the fact that the dislocations which these generators produce will not be able to cut across the slip lamellas associated with the active system unless the applied stress for the latent planes is as large as that required to induce continued flow in the active planes, or, if an equation of the type (7) is generally valid, unless the stress in the latent planes is closely related to that in the active planes. A theory based upon this viewpoint would fit

in with the possibility, discussed in § 6, that a given slip generator becomes inoperative after it has begun to generate dislocations if the dislocations become stuck and exert a reactive force on the generator that is sufficient to compensate for the force exerted by the applied stress. We might expect the force required to push a dislocation that moves in a latent plane obliquely across one of the slip planes of the active system, which is heavily studded with vacant lattice sites, to be closely related to that required to push a dislocation through the slip plane, since the forces exerted by vacancies should be primarily short in range and nearly isotropic. Once a dislocation which cuts obliquely across a slip plane becomes stuck in the latter, the vacancies near to it will diffuse toward it and bind it tightly so that a relative high stress will be needed to release it again.

On the other hand, if the generators for all slip planes are closely correlated in space, for reasons which are not apparent at the present time, the factors which make it necessary to increase the applied stress in order to induce generators for the active planes to form new dislocations could presumably have an influence on generators for the latent planes. It seems fruitless to speculate on this issue until we have a much more complete picture of the density of slip bands in actual materials.

C. Prismatic Slip

It is interesting to observe that the right-hand side of (7) can vanish under appropriate conditions. Consider, for example, the case of the thallium halides, which have been studied extensively by Smakula and Klein⁽⁶⁰⁾. The slip directions are the three (100) directions, whereas the slip planes are the six planes of the (110) type. The right-hand side of (7) will vanish in a case in which the latent slip system is chosen to have the same slip direction as the active one, but in which the latent and actual slip planes are orthogonal (*e.g.* the (110) and (1 $\bar{1}$ 0) planes, associated with the (001) directions). It would be interesting to know if the slip in the actual slip plane has a negligible effect on the latent plane in this case. It is possible that slip will occur on both systems as a result of inhomogeneous loading unless great care is taken, as apparently was done in the measurements of Kochendörfer and Röhm.

§ 9. CREEP IN PURE METALS⁽⁶¹⁾

If we postulate that dislocations generate vacancies and that clusters of these may entrain dislocations, somewhat in the manner in which carbon atoms entrain dislocations in steels, according to Cottrell⁽⁶²⁾, we should expect to achieve a gradual type of plastic flow at temperatures sufficiently high that diffusion of the vacancies or clusters may occur, so that they migrate with the dislocations. It is possible that the steady state of creep in pure metals occurs by this mechanism. In this case we should expect the activation energy associated with the steady-state creep process to be closely related to the activation energy for migration of vacancies or

clusters of vacancies. The value of interest may be different from that for the perfect crystal, since the diffusing vacancies or clusters are near dislocations. On the other hand, if the material contains sufficient impurity that the impurity atoms impede the motion of the dislocations by aggregation at the dislocations, we might expect the activation energy limiting creep to be determined by that of the diffusion coefficient of the impurity atom. If the latter is able to migrate by means of the Johnson-Wagner mechanism, the vacancies generated during motion of the dislocations may be present in sufficient numbers that those produced thermally are relatively unimportant. In this case the interesting component of the activation energy of the impurity atom would be that associated with the jump frequency of the coupled vacancy and impurity atom. This activation energy may be significantly different from that for migration of isolated vacancies or clusters of vacancies.

It is possible that at elevated temperatures, where the vacancies are highly mobile, those generated by the dislocations succeed in transporting to the dislocation foreign atoms which would otherwise be unable to migrate. In this case the vacancies and impurity atoms combine to produce a Cottrell-type impediment which would not be effective if either imperfection were present alone. Similarly, it is possible that at sufficiently low temperatures, where the vacancies would form impeding clusters if uninhibited, they may become attached to impurity atoms instead and become frozen fast because the latter are not able to diffuse rapidly enough to aggregate at the dislocations, even with the aid of the Johnson-Wagner mechanism. In this case the impurities would have the effect of raising the transient creep rate by inhibiting the vacancies from forming clusters⁽⁶³⁾.

There is considerable evidence⁽⁶¹⁾ to show that the rate of steady-state creep varies with temperature through a Boltzmann factor. In fact, this is known as Ludwik's law. Moreover, this work shows that the activation energy appearing in the Boltzmann factor usually depends on the chemical composition and other factors, such as the history of the specimen, as if the critical diffusing entity might vary from one specimen to another.

There has been careful work on the creep of single crystals. For example, Boas and Schmid⁽⁶⁴⁾ have studied the creep of single crystals of zinc in the range of small strain where the transient and steady-state creep occur simultaneously and have found that the activation energy for this range is about 16 800 cal per mol. This value is to be compared with the activation energy for self diffusion obtained by Banks and Miller⁽⁶⁵⁾ parallel and perpendicular to the hexagonal axis, namely 20 400 and 31 000 cal per mol. Perhaps the migration of vacancies is the limiting process in this case. On the other hand, Cottrell and Aytakin⁽⁵⁶⁾ investigated the rate of steady-state creep of single crystals of zinc and came to the conclusion that the activation energy for the process is between 28 000 and 30 000 cal per mol, although the value for polycrystalline

zinc is between 11 000 and 13 000 cal per mol. The origin of this discrepancy between the measurements on single crystals is not known. Since both investigators employed specimens of comparable size and used comparable loads, the difference presumably arises either from a difference in purity of the two types of specimen, from a difference in the range of creep investigated, or from the difference in temperatures emphasized by the two groups of investigators. Boas and Schmid employed temperatures in the range from room temperature to that of liquid nitrogen, whereas Cottrell and Aytakin worked between room temperature and 125° C.

Dushman⁽⁶⁶⁾, who has made extensive creep measurements on a number of metals, and has tabulated results of other investigators in a form that is of interest for our purposes, gives the following activation energies :

TABLE I.

Activation Energies for Creep (Principally from Dushman).

Metal	Q (cal per gram atom)
Aluminium (pure)	14 100
Aluminium (commercial with Fe, Si, Mg)	33 140
Aluminium with 2% Mg	9 500
Lead (extruded rod)	3 000
Lead (polycrystalline)	11 500
Platinum	42 100
Silver (vacuum annealed 550° C)	12 000
Silver (air annealed 550° C)	19 100
Silver (annealed in hydrogen 550° C)	25 000
Tin (polycrystalline)	5 290
Zinc	$\left\{ \begin{array}{l} 10\,600 \\ 29\,000-30\,000 \end{array} \right.$

Although the activation energy for creep in "pure" metals, such as aluminium and the specimens of zinc employed by Boas and Schmid, appears to be in about the range that one might expect if the diffusion of vacancies is the essential limiting factor, the values for other pure metals, such as lead (extruded rod) and silver that has been vacuum annealed seem to be much on the low side. The activation energies for self diffusion in the latter metals⁽⁶⁷⁾ are 27 000 and 42 000 cal per mol, respectively. However, the density of impurity atoms in any of the foregoing materials is probably considerably higher than the density of vacancies generated during the flow process, so that the impurities probably condition the flow rate even in the metals which are considered pure. As we saw above, impurities might shift the activation energy in either direction. It would appear to be highly desirable to study creep in metals which have an intrinsic purity analogous to that of the best salts, namely of the order of one part per million of impurity. Selected impurities might then be added under highly controlled conditions, much in the way that impurity atoms are added to semi-conductors to obtain highly controlled conditions.

It is interesting to note that magnesium decreases the activation energy for creep in aluminium. This may be an example of a case in which the impurity, which has a larger diameter than the solvent, captures the vacancies formed during creep and prevents them from forming aggregates at the dislocation. On the other hand, the impure aluminium containing iron and silicon, as well as magnesium, has a higher activation energy than the "pure" aluminium. This may be a case in which the vacancies assist the impurities to migrate to the dislocation and impede its motion.

§ 10. CHANGE IN DENSITY DURING COLD-WORK

It has been recognized⁽⁶⁸⁾ for many years that the density of metals is reduced by cold-work. Values of the decrease in the range of 0.1% have apparently been reported in the literature in cases of heavy cold-work. In fact, measurements on brass by Masima and Sachs⁽¹⁰⁾ were discussed in § 2, C. Two factors may contribute to decrease the density: (1) Vacancies and interstitial atoms may affect the density provided they are generated in the manner proposed here. If vacancies are generated preferentially, as we have supposed for the sake of simplicity, we should expect the fractional decrease to be closely equal to the fraction of vacant sites. If vacancies and interstitial atoms are generated in nearly equal numbers, we should expect the influence of the vacancies in reducing the density to dominate over the influence of the interstitial atoms in increasing the density because the latter will not fit nearly into interstitial places, but will have an expansive influence on the lattice. (2) Dislocations may also be expected to expand⁽⁶⁹⁾ the lattice and reduce the density. A Taylor dislocation contains both compressive and tensile regions. Since the interatomic forces resisting strain oppose compressive strain more effectively than tensile strain, we should expect the dislocations to have the net effect of reducing the density.

We saw in § 2 that a fraction of approximately 2×10^{-5} vacancies are introduced in copper as a result of a 10% strain. Hence a strain of 5 should be sufficient to decrease the density by about 0.1% as a result of the production of vacancies alone, provided there are no compensating effects. In fact, measurements of density may provide a very accurate method of determining the density of vacancies produced during cold-work if it proves possible to establish the fact that vacancies are the predominant imperfection, as may well be the case.

The density change arising from vacancies should subside at lower temperatures than some of the other effects which attend cold-work, such as work-hardening. For the vacancies may cluster into two-dimensional plates which collapse to form small dislocation rings. These rings may impede the migration of dislocations, and hence contribute to work-hardening; however, their contribution to the decrease in density should be small compared to that which the vacancies would exert if fully dispersed.

§ 11. COMPARISON OF THE EFFECTS OF COLD-WORK WITH THOSE OF BOMBARDMENT⁽⁷⁰⁾

It has been well established⁽⁷¹⁾ that atoms may be displaced from normal lattice positions as a result of bombardment with massive particles such as alpha rays or the particle radiations produced by accelerators. The imperfections produced by bombardment should consist of interstitial atoms and vacant lattice sites in pure metals which have been irradiated at temperatures sufficiently low that annealing will not occur as a result of diffusion. It should be relatively easy to duplicate the densities of imperfections engendered by moderate amounts of cold-work in this way. It does not follow, however, that the change in physical properties produced by bombardment will duplicate those which accompany cold-work. For, as we have seen in previous sections, it is possible that vacancies are the dominant imperfection formed by cold-work and that their effect on properties will be substantially different from that of other imperfections, such as interstitial atoms. For example, it is possible that the reduction of density of the specimen resulting from N vacant sites per unit volume is much greater than the reduction which occurs when there are $N/2$ vacancies and $N/2$ interstitial atoms per unit volume. On the other hand, the change in electrical resistivity and Young's modulus associated with the second type of imperfection may be much greater than that associated with the first.

In fact a comparison of the properties of irradiated materials and those which have been cold-worked may furnish a means of deciding whether or not vacancies are the dominant imperfection in cold-worked materials.

§ 12. PROPOSED EXPERIMENTS

It is fairly evident that many of the crucial decisions concerning the explanations offered in the preceding sections will have to be based on appropriately chosen experiments. The writer would like to propose the following experiments for investigation.

(1) It would be highly desirable to have the experiments of Gyulai and Hartly repeated under circumstances which permit a more definite evaluation of the density of vacancies produced during the cold-working process. In particular it would be valuable to know the absolute magnitude of the conductivity immediately after plastic flow. If the technique for producing pure shear flow described by Röhms and Kochendörfer⁽⁴⁸⁾ can be applied, it may prove possible to make measurements during the flow process.

(2) Although the experiments on the change of electrical resistivity of metals with cold-work are highly suggestive, it would be desirable to extend the range of investigation in several ways: (a) The work should be extended to single crystals; (b) the total strain could be increased by rolling, as in the experiments of Gayler described in § 5; (c) simultaneous

X-ray and density measurements could be made to see how closely the changes in properties can be correlated with the production of vacancies : (d) the influence of specific impurity atoms on the rate of recovery could be studied to see if vacancies may be retained in unaggregated form by the impurities, as the somewhat rough experiments of Tammann and his associates indicate is the case. In this connection it would be interesting to know if the rate of diffusion of a foreign atom or a radioactive tracer can be increased by continuous plastic flow, for example, by cyclical bending at an elevated temperature where work-hardening is avoided.

(3) Experiments on the energy stored during cold-work evidently provide an important key to determining the importance of the vacancies or interstitial atoms during cold-work. It would be very valuable to know to what extent this release runs parallel to the recovery of the increase in electrical resistivity and to the resoftening of the metal. It would also be interesting to know if appropriate additions of impurity atoms, such as the addition of magnesium to aluminium, can aid in decreasing the rate of release of an appreciable part of the stored energy, as might be expected if foreign atoms can seize the vacancies and prevent them from recombining. Shuttleworth has pointed out to the writer that it may prove possible to make significant calorimetric measurements of this type below room temperature by choosing the metals under investigation properly. In this case it would be far easier to attain precision than at elevated temperatures where radiation losses and similar effects are important.

(4) The experiments on aluminium alloys containing copper, described in § 5, seem to show that copper atoms can capture vacant sites in this system. It would be very valuable to study the properties of such alloys further in the light of the viewpoint presented here.

(5) The experiments of Masima and Sachs⁽¹¹⁾ on the generation and annealing of density changes as a result of plastic flow in brass should be extended to simpler metals and extended to lower temperatures.

(6) The experiments of Heidenreich and Shockley, Brown, Craig and Czyzewski represent merely a start in the problem of determining the properties of slip bands. It would be very valuable to know if the phenomena observed in aluminium and brass occur generally in other face-centred cubic metals and in metals or salts having other structures. Specific measurements that might be carried out with aluminium or brass is as follows : (a) It would be valuable to know if further slip can occur along a given slip line if the specimen is resoftened after flow, at the lowest possible temperature without recrystallization. An experiment of this type might determine whether the generator is destroyed incidental to the production of 1000 dislocation rings, or whether it is only temporarily incapacitated because of the distortion of the slip plane on which its dislocations move. (b) It would be interesting to produce fiducial markers on a specimen by permitting a small amount of slip in one family of slip planes, sufficient to form a few slip lines, to anneal the specimen so that

it is again soft, and then allow slip in another family of slip planes. Investigation of the change in orientation, if any, of the first set of lines might allow one to decide whether all the flow occurring in the second set of slip planes is associated with slip bands or whether an appreciable component occurs between the observed slip lines.

(7) It would be very valuable to carry out creep measurements in cases in which the information derived from experiments of the type (2) is available, in order to see if there is a correlation between the activation energy for creep and that for recovery of electrical conductivity or other properties which are assumed to change because of migration of vacancies. In cases in which creep experiments indicate that the vacancies produced during creep aid impurity atoms to diffuse to the dislocation and produce locking stresses of the type considered by Cottrell, it would be valuable to know if the impurities have the effect of introducing upper yield-point phenomena. As is suggested in the text (§ 8), specimens which possess no more than one part per million of uncontrolled impurity would be highly desirable. The properties of this base material could then be altered systematically by the addition of known impurities, in analogy with the very profitable investigations which have been carried out on the electrical properties of non-metals, such as the alkali halides and silicon.

(8) It would be valuable to make a systematic comparison of the changes in physical properties, such as electrical resistivity, density and elastic constants, produced by cold-work and by bombardment with heavy particles in order to see if the two are identical or if there is a qualitative difference, as might be expected if cold-work produces vacancies in preference to interstitial atoms.

The foregoing manuscript was drafted first early in the winter of 1950-51. The writer would like to express his gratitude to a number of colleagues who have made critical comments. He is particularly grateful to Professors H. Brooks and N. F. Mott and to Dr. L. Slifkin.

REFERENCES

- (1) F. SEITZ, 1950, *Phys. Rev.*, **80**, 239.
- (2) Z. GYULAI and D. HARTLY, 1928, *Zeits. f. Phys.*, **51**, 378.
- (3) A. W. STEPANOW, 1933, *Zeits. f. Phys.*, **81**, 560.
- (4) J. MOLENAAR and W. H. AARTS, 1950, *Nature, Lond.*, **166**, 690.
- (5) J. O. LINDE, 1950, *Ann. d. Phys.*, **15**, 219.
- (6) See, for example, the discussion by the writer, 1950 *Acta, Crystallographica*, **3**, 355.
- (7) G. I. TAYLOR and H. QUINNEY, 1934, *Proc. Roy. Soc.*, **143**, 307; *Ibid.*, 1937, **163**, 157.
- (8) See, for example, H. W. ETZEL and R. J. MAURER, 1950, *Journ. Chem. Phys.*, **18**, 1003.
- (9) See note added in proof, reference 1.
- (10) C. WAGNER and P. HANTELMAUN, 1950, *Journ. Chem. Phys.*, **18**, 72.
- (11) M. MASIMA and G. SACHS, 1928, *Zeits. f. Phys.*, **51**, 321; *Ibid.*, 1929, **54**, 666.

- (12) G. TAMMANN and K. L. DREYER, 1933, *Ann. d. Phys.*, **16**, 111; G. TAMMANN and G. BANDEL, *Ibid.*, p. 120; G. TAMMANN and K. L. DREYER, *Ibid.*, p. 657; G. TAMMANN and G. MORITZ, *Ibid.*, p. 667; G. TAMMANN and V. CAGLIOTI, *Ibid.*, p. 680. This work includes measurements of the change in thermoelectric potential as a result of cold-work. Transient effects resemble somewhat those observed by Gyulai and Hartly are described for a number of metals and alloys. I am indebted to Professor N. F. Mott for calling my attention to these papers.
- (13) The possibility that dislocations move at a velocity close to that of sound was first exploited by F. C. FRANK, 1948, *Bristol Conference on the Strengths of Solids*, p. 46 (London: Physical Society), who attempted to employ this property to achieve dynamic multiplication of dislocations. The dynamics of such motion has been investigated by J. D. ESHELBY, 1949, *Proc. Phys. Soc. A*, **62**, 307, under the assumption that the medium is continuous. Eshelby has demonstrated in a very elegant way that the energy varies with velocity in a manner resembling the variation of energy of a particle in the special theory of relativity.
- (14) G. LEIBFRIED, 1949, *Zeits. f. Phys.*, **127**, 344; see also F. R. N. NABARRO, 1950, *Phil. Mag.*, **41**, 1270.
- (15) F. SEITZ, 1950, *Pittsburgh Symposium of the Plastic Deformation of Crystalline Solids*, p. 1 (Naval Research Laboratory).
- (16) For a review of dislocation theory see, for example, *Bristol Conference on the Strength of Solids* (reference 13); *Pittsburgh Symposium on the Plastic Deformation of Crystalline Solids* (reference 15); A. H. COTTRELL, 1949, *Progress in Metal Physics* (New York: Interscience Press); W. T. READ and W. SHOCKLEY, 1950, *Phys. Rev.*, **78**, 275.
- (17) This possibility was suggested independently by W. T. READ (private communication), on the basis of a viewpoint developed by Read and Shockley and is to be published in the reports of the Pocono Conference on Imperfect Crystals.
- (18) See, for example, the paper by the writer, 1949, *Discussions of the Faraday Society*, No. 5, 271; and that by S. SIEGEL, 1949, *Phys. Rev.*, **75**, 1823.
- (19) H. B. HUNTINGTON and F. SEITZ, 1942, *Phys. Rev.*, **61**, 315; H. B. HUNTINGTON, 1942, *Ibid.*, **61**, 325.
- (20) F. SEITZ and T. A. READ, 1941, *Journ. App. Phys.*, **12**, 182; A. E. H. LOVE, 1927, *Mathematical Theory of Elasticity* (Cambridge: University Press).
- (21) C. TUBANDT, *Handbuch der Experimental Physik*, Vol. XII, part 1.
- (22) See, for example, the review by the writer, 1946, *Rev. Mod. Phys.*, **18**, 384.
- (23) H. W. ETZEL and R. J. MAURER, 1950, *Journ. Chem. Phys.*, **18**, 1003; D. E. MAPOTHER, H. N. CROOKS and R. J. MAURER, 1950, *Ibid.*, **18**, 1231.
- (24) H. KELTING and H. WITT, 1949, *Zeits. f. Phys.*, **126**, 697.
- (25) J. TELTOW, 1949, *Ann. d. Phys.*, **5**, 63, 71; 1950, *Zeits. Phys. Chem.*, **195**, 197, 213.
- (26) J. R. REITZ and J. L. GAMMEL (to appear in *Journ. Chem. Phys.*).
- (27) G. J. DIENES, 1948, *Journ. Chem. Phys.*, **16**, 620.
- (28) The writer accepts the measured value of M. S. MAIER and H. R. NELSON, 1942, *Trans. A.I.M.E.*, **147**, 39, as the most reliable one (see reference 8).
- (29) The writer is indebted to N. F. Mott for calling his attention to this aspect of the influence of cold-work on the properties of metals.
- (30) M. L. V. GAYLER, 1946, *Journ. Instit. of Metals*, **72**, 243, 543. See also the review by G. C. SMITH, 1949, *Progress in Metal Physics* (Butterworth Scientific Publications), and 1939, *Age Hardening of Metals* (Cleveland: American Society for Metals).

- (31) R. P. JOHNSON, 1939, *Phys. Rev.*, **56**, 814 ; C. WAGNER, 1937, *Zeits. Phys. Chem.*, **B**, **38**, 325.
- (32) See, for example, C. J. SMITHELLS, 1949, *Metals Reference Book* (London : Butterworth Scientific Publications) ; R. F. MEHL, F. N. RHINES and K. A. VON DEN STEINEN, 1941, *Metals and Alloys*, **13**, 41 ; H. BÜCKEL, 1943, *Zeits. f. Elek.*, **49**, 238 ; A. H. BIERWALD, 1939, *Zeits. f. Elek.*, **45**, 789.
- (33) F. R. N. NABARRO, 1949, *Symposium on Internal Stresses in Metals* (London : Institute of Metals), p. 237.
- (34) F. ROHNER, 1947, *Journ. Inst. Metals*, **73**, 285.
- (35) R. D. HEIDENREICH and W. SHOCKLEY, *Conference on the Strength of Solids* (reference 13), p. 57.
- (36) A. F. BROWN, 1949, *Nature, Lond.*, **163**, 961.
- (37) F. C. FRANK and W. T. READ, 1950, *Phys. Rev.*, **79**, 722.
- (38) This decrease is indicated in the sketches of Heidenreich and Shockley, which represent averages of their observations. Mr. W. J. Craig and Professor H. Czyzewski of the University of Illinois have carried out experiments similar to those of Heidenreich and Shockley and Brown on 70-30 brass and other metals. Their results on brass indicate that slip occurs in a manner very similar to that in aluminium. However, the spacing between lines in a band is about 450 Å. Moreover, the decrease in displacement on successive lines in a band is strongly marked.
- (39) H. J. GOUGH and H. L. COX, 1931, *Journ. Inst. Metals*, **45**, 71, have presented evidence to show that the spacing between slip bands is random in silver.
- (40) J. N. KELLAR, P. B. HIRSH and J. S. THORP, 1950, *Nature, Lond.*, **165**, 554.
- (41) R. W. CAHN, 1949, *Journ. Inst. Metals*, **76**, 121.
- (42) See the excellent review of the topic by W. T. Read. "Experimental Information on Slip Bands" (Pocono Conference of National Research Council, 1950).
- (43) F. SEITZ, 1950, *Phys. Rev.*, **79**, 723, 890, 1002, 1003 ; see also discussions of reference 15. The spiral prismatic dislocation appears to have four important characteristics : (1) It represents a highly stable form of a vacancy precipitate, analogous to the precipitate formed by aggregation of insoluble atoms in a solid or liquid, although potentially capable of achieving a much greater degree of registry with the surrounding lattice ; (2) It is a source of distortion of the type which appears to be associated with mosaic structure. Since the spiral dislocations possess a low energy, this distortion is very difficult to anneal out of the crystal. (3) It is a source of equi-spaced generators of the Frank-Read type, such as appears to be necessary to explain the equi-spaced slip lines in a slip band. (4) It may act as a weak spot in the lattice and provide points where rupture can occur for low tensile stress. It is evident from fig. 5 that an unclosed or partly closed cone of vacancies will possess a high stress magnification factor.
- (44) The viewpoint that the dislocations responsible for plastic flow are those having lowest energy has been expressed independently by B. Chalmers (private communication) on the basis of evidence to be presented in a forthcoming letter in *Nature*.
- (45) F. C. FRANK, 1949, *Proc. Phys. Soc. A*, **62**, 202.
- (46) See the papers of reference 16, particularly that of Mott and Nabarro in *Report of Conference of the Strength of Solids*, p. 1.
- (47) A. H. COTTRELL, *Pittsburgh Symposium on Plastic Deformation of Crystalline Solids* (reference 16), p. 60.

- (48) See the paper by U. DEHLINGER, *Pittsburgh Symposium on Plastic Deformation of Crystalline Solids* (reference 16), p. 103.
- (49) See, for example, C. F. ELAM, 1936, *The Distortion of Metal Crystals* (Oxford : University Press), for a review of this work.
- (50) B. L. AVERBACH, and B. E. WARREN, 1949, *Journ. App. Phys.*, **20**, 1066.
- (51) This explanation of the limited life of the generator was proposed independently by Read and Shockley (as yet unpublished paper in Pocono Conference of National Research Council).
- (52) O. HAASE and E. SCHMID, 1925, *Zeits. f. Phys.*, **33**, 416.
- (53) J. S. KOEHLER and F. SEITZ, 1947, *Journ. App. Mech.*, **14**, 217.
- (54) A. H. COTTRELL, *Report of Conference on the Strength of Solids* (reference 16), p. 30.
- (55) W. F. BROWN, 1940, *Phys. Rev.*, **58**, 730 : A. R. KAUFMANN, 1940, *Ibid.*, **57**, 1089.
- (56) T. A. READ and F. SEITZ, 1941, *Journ. App. Phys.*, **12**, 179.
- (57) A. H. COTTRELL and V. AYTEKIN, 1950, *Journ. Inst. of Metals*, **77**, 389.
- (58) T. E. TEITZ, R. A. ANDERSON and J. E. DORN, 1949, *Journ. of Metals*, **1**, 921 ; T. V. CHERIAN, P. PIETROKOWSKY, and J. E. DORN, 1949, *Ibid.*, **1**, 957.
- (59) See the book by C. F. ELAM, reference 49. Also the comments by T. A. READ and F. SEITZ, 1941, *Journ. App. Phys.*, **14**, 184 : R. D. HEIDENREICH and W. SHOCKLEY, reference 35.
- (60) A. SMAKULA and M. W. KLEIN, 1949, *Journ. Opt. Soc. Am.*, **39**, 445.
- (61) For a review of creep in relatively pure materials see the following articles : H. J. TAPSELL, 1931, *Creep* (Oxford : University Press) ; F. SEITZ and T. A. READ (reference 20) : F. SEITZ, 1943, *The Physics of Metals* (New York : McGraw-Hill Book Co.), p. 133 ; W. KAUFMANN, 1941, *Trans. A.I.M.E.*, **143**, 57.
- (62) A. H. COTTRELL, 1948, *Bristol Conference on the Strength of Solids* (London : Physical Society), p. 30.
- (63) Mott proposed this mechanism to the writer as an explanation of the fact that impurity atoms may have the effect of enhancing the steady-state creep rate at elevated temperatures.
- (64) W. BOAS and E. SCHMID, 1936, *Zeits. f. Phys.*, **100**, 463.
- (65) F. R. BANKS, 1941, *Phys. Rev.*, **59**, 376 : F. R. BANKS and P. H. MILLER, 1942, *Ibid.*, **61**, 648.
- (66) S. DUSHMAN, L. W. DUNBAR and H. HUTTENSTEINER, 1944, *Journ. App. Phys.*, **15**, 108.
- (67) G. VON HEVESY, W. SEITH and A. KEIL, 1932, *Zeits. f. Phys.*, **79**, 197 ; W. A. JOHNSON, 1941, *Trans. A.I.M.E.*, **143**, 107 : 1942, *Ibid.*, **147**, 331.
- (68) See, for example, the discussion by C. ZENER, 1942, *Trans. A.I.M.E.*, **147**, 361 ; also the table in E. SCHMID and W. BOAS, 1950, *The Plasticity of Crystals* (London : F. A. Hughes and Co.), p. 202.
- (69) See the discussion by the writer following the paper of reference 68, p. 368.
- (70) The writer is indebted to S. Siegel for a valuable discussion of this topic.
- (71) See, for example, the review by the writer in reference 18 as well as the papers by Siegel and the group at Purdue University. See also the paper by W. H. BRATTAIN and G. L. PEARSON, 1950, *Phys. Rev.*, **80**, 846.

Crystal Growth and Dislocations

By F. C. FRANK

H. H. Wills Physical Laboratory, University of Bristol

THE THEORY OF CRYSTAL GROWTH

§1. INTRODUCTION

THE history of the theory of crystal growth divides into two parts. One is the theory of the growth of ideally perfect crystals, starting with Willard Gibbs (1878), and then developed between 1920 and 1948 by various workers, notably Volmer (1920 onwards), Kossel, Stranski, Becker and Döring, Frenkel, and Burton and Cabrera. The second is the theory of the growth of imperfect crystals, commencing with Burton, Cabrera and Frank (1949). The second makes use of the results of the first: basic to both is some understanding of the atomic nature of a crystal surface in equilibrium. In the early work this point is disposed of by tacit assumption, and its theoretical study has only recently been undertaken (Burton and Cabrera 1949, 1951).

§2. THE EQUILIBRIUM STATE OF A CRYSTAL SURFACE

The usual tacit assumption, which is now somewhat better founded, is that the faces of low index of a perfect crystal in equilibrium with its vapour or solution are fundamentally flat. The situation is represented in fig. 1, which represents a complete flat surface partially covered by another surface. A new layer is built up by adding molecules to the 'exchange' or 'Kossel' sites A, and the energy gained on adding a molecule to such a site is equal to the evaporation energy W . As the temperature is raised the surface will change in two ways:

(i) A small number of adsorbed molecules, marked B in fig. 1 (*b*), will appear on the surface. The proportion n_s of surface sites occupied by such molecules will be given approximately by

$$n_s = \exp(-W_s/kT) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where W_s is the energy required to transfer a molecule from a site A to a site B. We estimate W_s to be about half* the binding energy W .

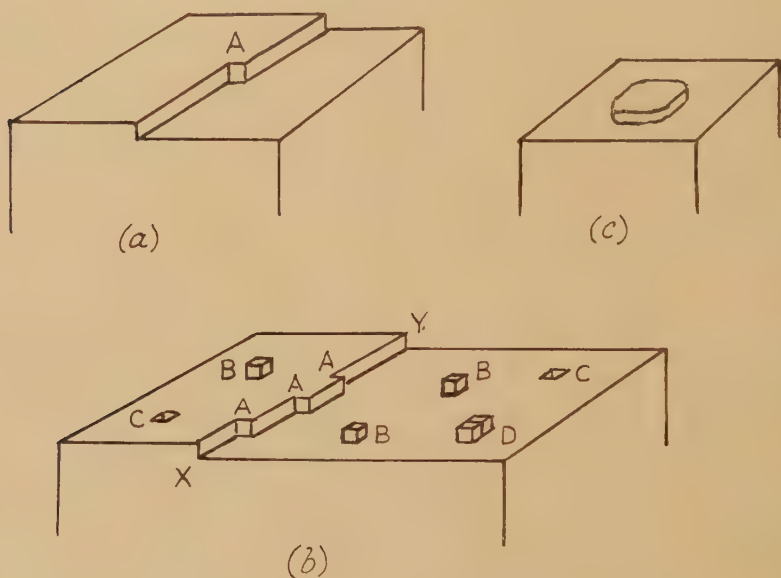
There is a similar number of surface vacancies, C, and there are smaller numbers of pairs of adsorbed molecules, D, or pairs of vacancies, and likewise of aggregates of larger numbers of molecules, the proportion of these being of progressively lower order of magnitude, by factors similar to n_s , as the number of molecules in the aggregate increases.

* Calculations by Mackenzie (Bristol thesis) for an interaction energy between molecules of the type $ar^{-12}-br^{-6}$ give for the energy of adsorption in B sites ($W-W_s$) about $0.6W$.

(ii) The number of exchange sites of type A increases. The relevant formula is given by eqns. (2) and (4) below. We have to emphasize, however, that in equilibrium the number of lines XY does *not* increase. The entropy gained in increasing this number does not compensate for the energy*.

The derivation of these results is simple so long as the singularities A, B, C, D, etc., are relatively rare departures from the low temperature configuration of fig. 1 (a); i.e. in a range of sufficiently low temperature. To determine what is a sufficiently low temperature, it is necessary to examine the problem in the whole temperature range, when one may no longer make approximations dependent upon n_s being a small number.

Fig. 1.



Schematic description of a crystal surface

- (a) at low temperatures ;
- (b) at higher temperatures ;
- (c) surface with an island.

The problem must then be treated as a 'co-operative' one. This has been undertaken by Burton and Cabrera (1951), who arrive at the conclusion that for every rational face of a crystal there is a critical temperature, above which its surface roughness on the atomic scale increases rapidly with increase of temperature. This may be called a surface melting temperature. These temperatures have the same order of magnitude as the ordinary melting point of the crystal, but are lower

* In the same way, the number of line dislocations in a crystal in thermal equilibrium is zero.

in the faces of higher index. It appears probable that most crystals retain some faces which have not 'melted' in this sense, up to the ordinary melting point, when they are in contact with their vapours. They probably also retain some 'unmelted' faces, in contact with their dilute saturated solutions in reasonably efficient solvents (since in such cases the temperatures as well as the surface energies are lower). This is probably not always the case when crystals are in contact with their own melts: such cases must be distinguished from the rest in the theory of crystal growth, and we shall return to their consideration later.

Since the 'melted' faces will grow rapidly and thus disappear, the problem of crystal growth is concerned with the unmelted faces, at any rate in vapours and solutions.

At all temperatures at which there exists, in the given environment, a set of 'unmelted' faces capable of enclosing a geometrical solid, any other possible face of the crystal may be described as one of these principal faces, modified by 'steps', one or more lattice spacing in height, of the types shown in fig. 1.

Steps of any orientation in a given face can in turn be described in terms of steps in certain principal directions in the face, modified by 'kinks', as at A in fig. 1 (a), which make a sideways displacement of the step by one lattice spacing. There is no critical temperature for the 'melting' of a monolayer step. As already stated, on raising the temperature more and more kinks (exchange sites) appear. As was shown by Frenkel (1945) and Burton and Cabrera (1949), if n_+ and n_- is the number of kinks of opposite sign per atomic spacing measured along the nearest principal direction, then in equilibrium

$$n_+n_- = \exp(-2w/kT), \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where w is the energy of formation of a kink. We estimate that for close-packed crystals with homopolar binding w is of the order of magnitude of 1/12 of the evaporation energy. The angle between the mean direction of the step and the nearest principal direction is given essentially by

$$\tan \theta = n_+ - n_-, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

so that the mean kink density, solving for n_+ and n_- from (2) and (3), is

$$n_+ + n_- = [4 \exp(-2w/kT) + \tan^2 \theta]^{1/2}. \quad . \quad . \quad . \quad . \quad (4)$$

This is usually quite a high density: the lowest temperature at which one is likely to grow a crystal from its vapour is, let us say, that at which its vapour pressure is 10^{-10} atmospheres. W/kT is then 23, and w/kT therefore about 2. Then, even in the closest-packed direction of a step, kinks occur every three or four atoms apart. Under the same conditions the fraction of surface lattice sites occupied by single molecules or surface vacancies is only about 10^{-5} . These considerations apply to crystals whose vapour is monatomic, or consists of molecules which have rotational freedom in the solid at its melting point. The kink density corresponding to a given vapour pressure will be smaller when the vapour

molecules possess rotational degrees of freedom not present in surface molecules of the solid: similarly, it will be higher for ionic crystals, but it will never be zero, or negligibly small.

One other important consideration about the state of the crystal surface first emphasized by Volmer is the ease of surface diffusion. The activation energy U_s for surface diffusion, i.e. migration from one surface site B to another is believed to be very low. Calculations by Mackenzie already referred to estimate that

$$U_s \sim \frac{1}{20} W.$$

For our theory an important quantity is the distance x_s diffused by a molecule on the surface before evaporating. For this purpose we introduce the adsorption energy W'_s at sites B, defined in terms of quantities already introduced by

$$W'_s = W - W_s.$$

Then the chance per unit time that an atom evaporates is

$$\nu \exp(-W'_s/kT).$$

The jump frequency is

$$\nu \exp(-U_s/kT),$$

so that the number of jumps is $\exp(W'_s - U_s)/kT$. To find x_s we take the square root and multiply by a ; thus

$$x_s = a \exp[(W'_s - U_s)/2kT]. \quad . \quad . \quad . \quad . \quad . \quad (5)$$

x_s/a is several hundred under the simple conditions considered above. It follows from this that in growth from a vapour the rate of direct arrival of molecules from the vapour at any particular point on a crystal surface is generally small compared with the rate of indirect arrival by way of surface migration.

The points where growth actually occurs are the kinks in the steps. The addition of one molecule at a kink constitutes Kossel's 'wiederholbare Schritt' of crystal growth. The total energy released in bringing a molecule from the vapour to this point is precisely the evaporation energy W . In equilibrium with the vapour, molecules join and leave these points with equal frequency. The rate of departure depends only on temperature, while the rate of arrival is proportional to vapour concentration (or more precisely, to the local concentration of adsorbed molecules). Hence in super saturated vapour more molecules join the kinks than leave them and so the step advances. The formation of kinks in a step is an easy process, the activation energy ($w + U_s$) being low, and we may assume as a good approximation that the kink density in a step of given orientation retains its equilibrium value whether growth is proceeding or not. We may not make this assumption for the step density; as already stated, in equilibrium there will be at most one in a principal face. On a perfect crystal the growth process removes all steps initially present to the boundaries of the principal faces, i.e. to the crystal edges.

§3. SURFACE NUCLEATION : THE THEORY OF GROWTH OF PERFECT CRYSTALS

Further growth of a perfect crystal, beyond the polyhedron of principal faces circumscribed to the initial shape, requires the initiation of new layers of the crystal. As was first appreciated by Gibbs (1878) this is a nucleation process, essentially like the nucleation of droplets from a vapour. Before a new layer can grow, as island monolayer (fig. 1 (c)) must form on the surface.

Any island monolayer of finite size must have a higher vapour pressure than the infinite crystal. This increase may be attributed to the specific free energy of the edge of the monolayer. There will thus be a critical linear size of nucleus r , for a given degree of supersaturation, for which the nucleus is in unstable equilibrium with the vapour. If the size of the nucleus is greater than r the nucleus will grow; if it is less it will evaporate.

If we assume the nucleus to be circular, elementary considerations give

$$r = a\phi / 2kT \ln \alpha, \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where α is the ratio of the pressure p to the equilibrium vapour pressure p_0 ($\alpha = p/p_0$), and $\frac{1}{2}\phi$ is the energy of the edge per atom. This is because the rate of evaporation from the curved edge is increased by

$$\exp(a\phi / 2rkT),$$

while the rate at which atoms arrive exceeds by the factor α the rate under equilibrium vapour pressure. Thus

$$\alpha = \exp(a\phi / 2rkT),$$

from which (6) follows. For more exact calculations we have to estimate the equilibrium shape of the nucleus. This has been done by Burton and Cabrera (1951) who find that the equilibrium shape is given by the formula

$$h(\theta) = s_0 \gamma(\theta) / kT \ln \alpha, \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where $h(\theta)$ is the distance from the origin of a tangent to the boundary of the island, which makes an angle θ with a fixed axis, s_0 is the area per molecule in the layer, and $\gamma(\theta)$ is the specific free energy of the boundary at orientation θ . This result is a generalization, in two dimensions, of Wulff's theorem, which relates the equilibrium shape of a body to its polar diagram of boundary free energy, and of the Gibbs-Thomson formula which relates its equilibrium size to the supersaturation. In particular, at points where a circle (of radius r) centred on the origin touches the boundary, we have

$$r = s_0 \gamma / kT \ln \alpha.$$

We shall refer to this, loosely, as the radius of the critical nucleus, and sometimes disregard the fact that it is not circular. It corresponds to (6)

if $s_0\gamma$ is replaced by $\frac{1}{2}a\phi$, where ϕ is an energy which in the simplest approximation is the neighbour-neighbour binding energy in the crystal—say about one-sixth of the evaporation energy per molecule.

Any smaller island than is specified by (6) has a much higher probability of evaporating than of growing larger. Any larger one has a much higher probability of growing than of shrinking. The island specified by (6) is referred to as the critical nucleus. The formation of a new layer thus depends on improbable fluctuations among aggregates of adsorbed molecules ultimately producing a critical nucleus. The probability of this occurrence is proportional to $\exp(-A_0/kT)$, where A_0 is half the total edge free energy of the critical nucleus. Apart from a numerical factor which can be calculated, and is close to one (it must for example lie between $\pi/4$ and 1 on a surface of fourfold symmetry), this 'activation energy for nucleation' is given by

$$A_0 = 2r\phi/a = \phi^2/kT \ln \alpha. \quad (8)$$

The rate of formation of critical nuclei is given by the formula

$$Z \cdot (S/s_0) \cdot \exp(-A_0/kT), \quad (9)$$

where S is the surface area of the crystal face under consideration and Z is essentially the rate of arrival of fresh molecules at single surface lattice sites. A formula of this type was first surmized by Volmer, and derived by Becker and Döring (1935) with the assumption of rectangular nuclei, within which restriction they could show that nuclei departing significantly from the square shape were unimportant.

Even in a dense environment, Z cannot exceed about 10^{13} sec^{-1} . On a crystal of millimetre dimensions, therefore, the whole factor outside the exponential must be less than 10^{27} : in fact it is commonly around 10^{22} . For an appreciable growth rate of 10^{-3} layers per sec (about 1 micron per month) it follows that $\ln \alpha$ must be at least $(\phi/kT)^2/90$, which with a typical value of (ϕ/kT) signifies a supersaturation not less than 25%. 50% or more is commonly necessary. The nucleation rate is thus an extremely sensitive function of supersaturation. For a small change of the latter, it changes from a quite negligible value, to one which imposes no significant hindrance to growth whatever. Moreover, quite high values are necessary for this to occur.

Volmer and Schultze (1931), thinking to verify this result, found a very different one. They studied the growth rate of individual iodine crystals in slightly supersaturated vapour at 0° C. For supersaturations above 1% the growth rate was a proportional to supersaturation. Only below this degree of supersaturation did it fall below proportionality, and then not abruptly. Their results are shown in fig. 2, and are discussed further later in this article.

On the other hand, the observations of Haward (1939) correspond strikingly to the expectations of surface nucleation theory. He studied the deposition of various sublimable solids on a metal surface, coated with a previous deposit of the same substance, held at a fixed temperature

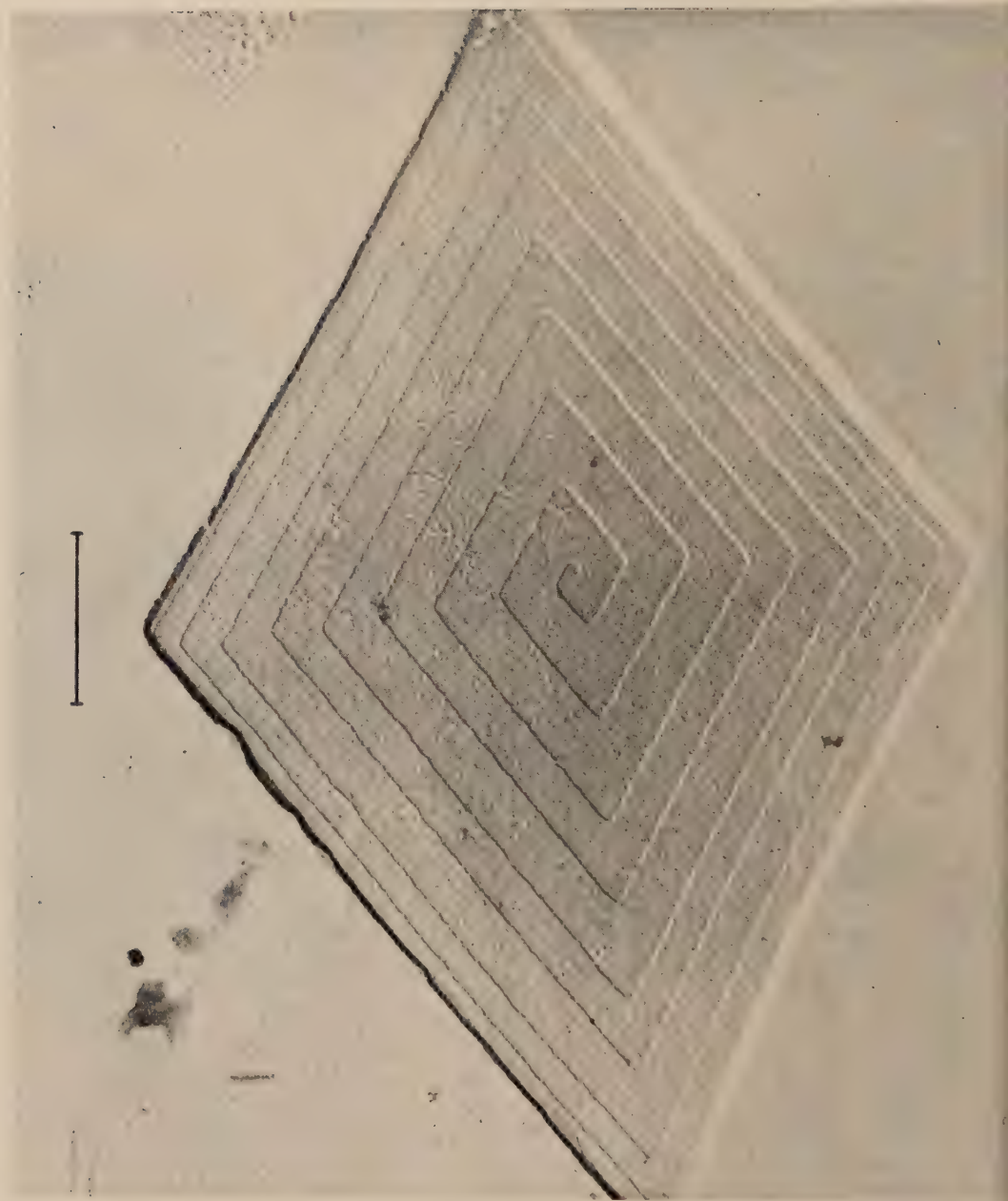
Fig. 5.



Growth steps on Beryl : $\text{Be}_3\text{Al}_2\text{Si}_6\text{O}_{18}$ (Griffin 1950).

$\times 1050$

Fig. 6.



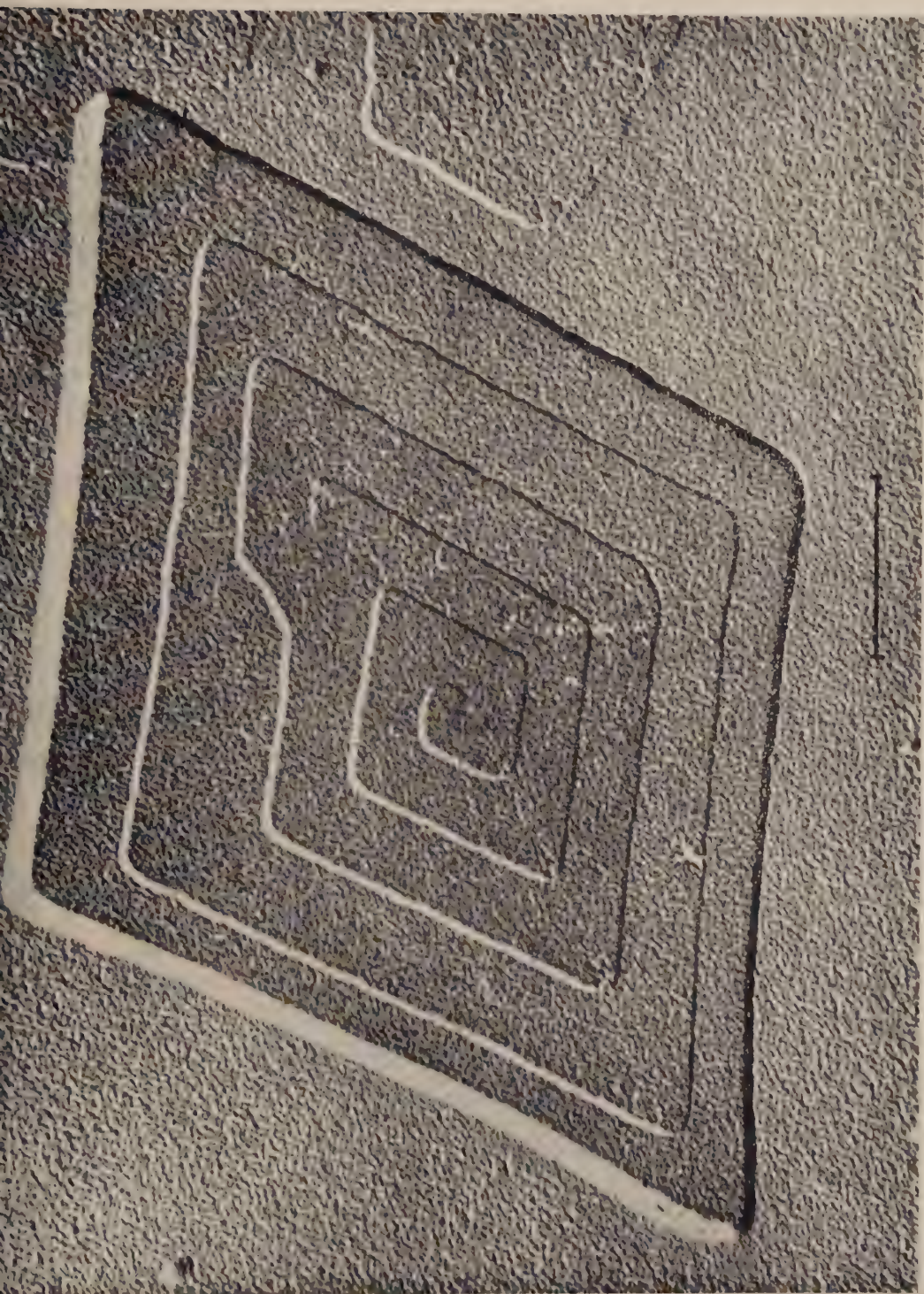
Paraffin n - $C_{36}H_{74}$ crystallized from petroleum ether on formvar substrate.

Gold-palladium shadowed.

$\times 2400$

Electron micrograph by H. F. Kay and Mrs. B. J. Appelbe after Dawson and Vand (1951).

Fig. 7.



Paraffin $n\text{-C}_{36}\text{H}_{74}$ crystallized from petroleum ether on formvar substrate.
Gold-palladium shadowed. $\times 27\,000$
Electron micrograph by H. F. Kay and Mrs. B. J. Appelbe after Dawson and Vand (1951).

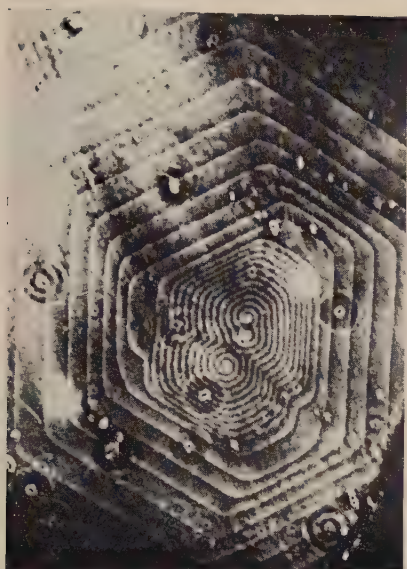
Fig. 8.



×200

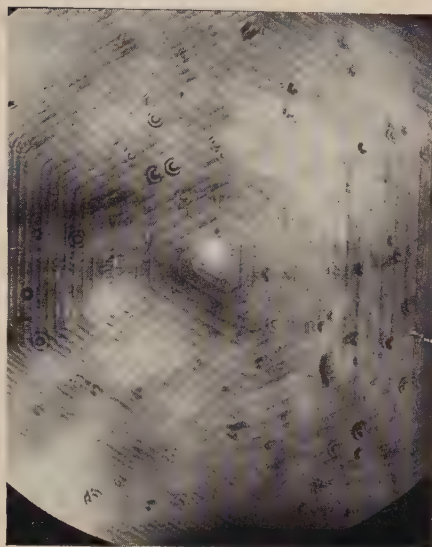
Growth steps on carborundum, SiC (Verma 1951).

Fig. 9.



$\times 280$

Fig. 10.



$\times 72$

Fig. 11.



$\times 132$

Figs. 9-11. Growth steps on carborundum, SiC (Verma 1951).

Fig. 12.



Fig. 13.

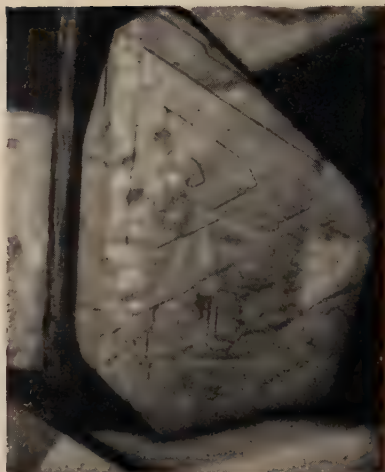


Fig. 14.

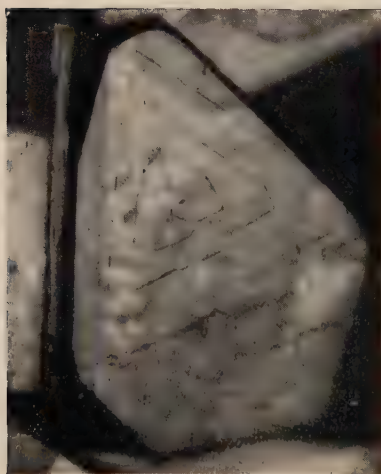


Fig. 15.



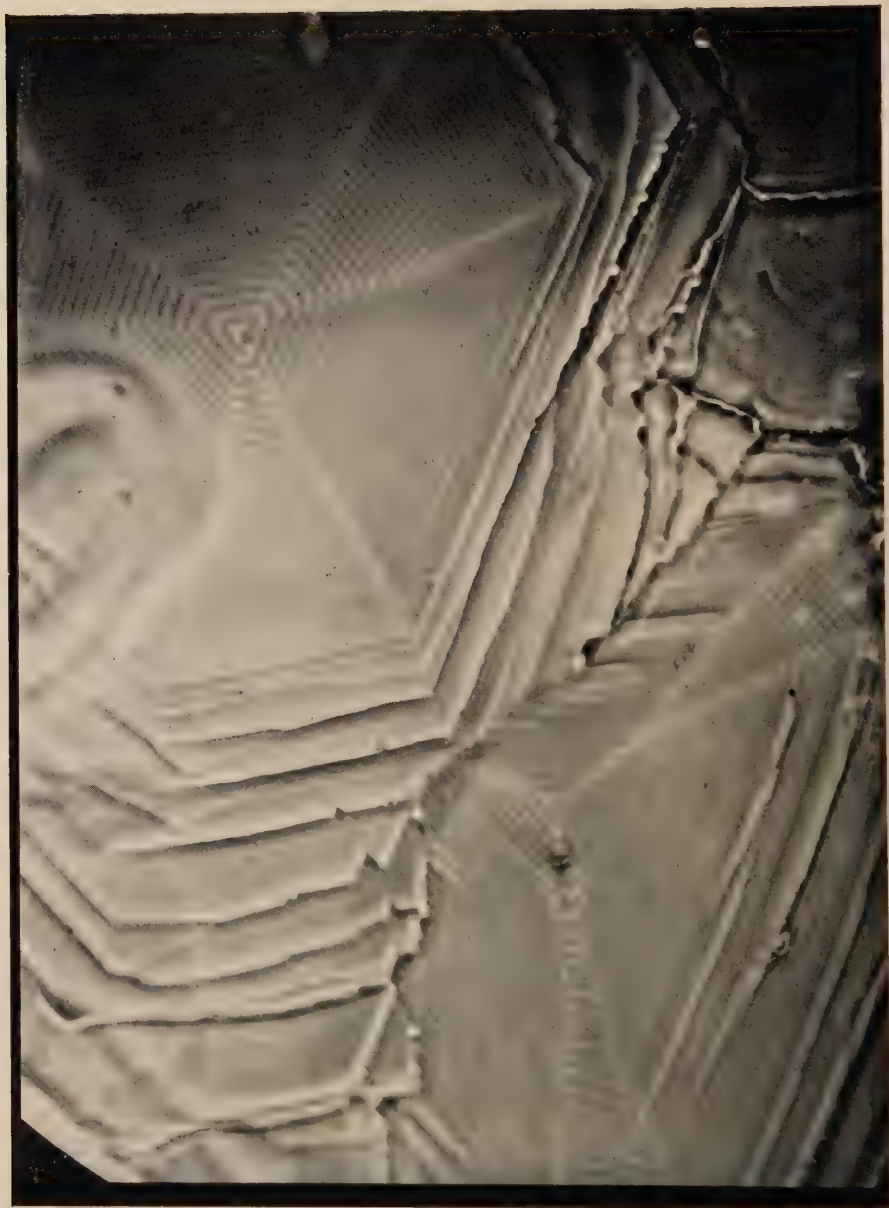
Figs. 12-15. Successive stages in the growth of cadmium iodide, CdI_2 , from aqueous solution (Forty 1951). $\times 350$

Fig. 16.



Internal interference micrograph of cadmium iodide crystal (Forty).
Step height (main system) $96 \pm 1 \text{ \AA}$, $\times 600$

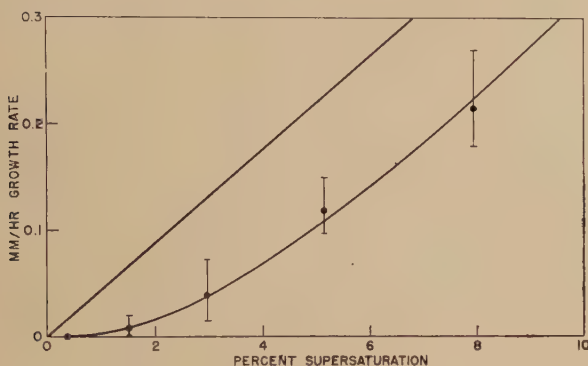
Fig. 17.



Phase contrast micrograph of dry cadmium iodide crystal (Forty).
Step heights: single centre $185 \pm 3 \text{ \AA}$; double centre $156 \pm 3 \text{ \AA}$. $\times 650$

and subjected to a vapour 'beam' of controlled intensity. The initial deposit was made by holding the target at a much lower temperature. He used a weighing technique which measured the mean overall rate of

Fig. 2.

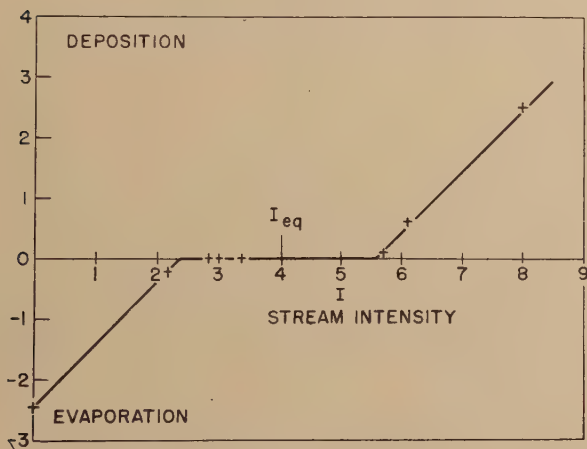


Growth rate of I_2 crystals from vapour at 0°C (Volmer and Schultze 1931). Mean value and range of scatter shown for 7 crystal faces (omitting one anomalously slow-growing face, 2-II).

Straight line: Hertz law, $\beta=0.4$.

Curve: eqn. (15), $\beta=0.4$, $\sigma_1=0.1$.

Fig. 3.



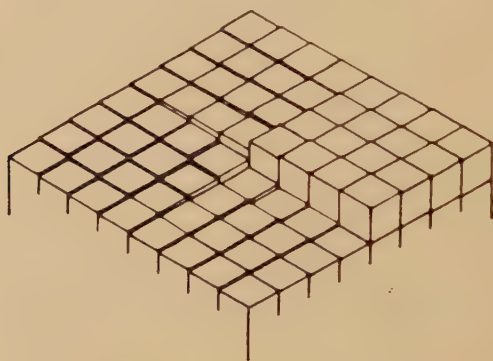
Rate of deposition, D (and evaporation, $-D$) at HgI_2 surface at 22.9°C , vs. vapour stream intensity I . I_{eq} is the rate at which molecules from saturated vapour would strike the surface. D and I are in the same arbitrary units. Plotted from data of Haward (1939).

deposition and would have been insensitive to the growth of a small proportion of crystals on the target. Fig. 3 is plotted from his data for HgI_2 at 22.9°C . As the figure shows, there is no appreciable growth

unless the beam intensity is greater than the rate of arrival of molecules from vapour at about 40% supersaturation, nor evaporation unless the intensity is less than the rate of arrival at 40% sub-saturation; and the rate of deposition or evaporation is equal to the excess or deficit from these critical values. (These figures are rough, since I_{eq} is an extrapolated figure, and also contains all the calibration errors of the apparatus. The ratio of the critical intensities is more reliable.) The ratio between the critical values increases slowly with decrease of temperature. The result is in astonishingly straightforward agreement with the theory of surface nucleation. It suggests, surprisingly, that most of the crystals in the compact deposit have perfect crystal surfaces, and are able in some way to protect each other's edges.

Similar observations were made with anthracene, the ratio of the two critical intensities being nearer to three.

Fig. 4.



The end of a screw dislocation in a building-block crystal model. Each cube represents a molecule.

There are no other observations of this kind known to the author. It is the general experience that *those crystals which do grow* do so at rates proportional to supersaturation down to supersaturations much lower than the theoretical critical supersaturation of surface nucleation theory.

THE GROWTH OF IMPERFECT CRYSTALS

§4. GENERAL THEORY

X-ray diffraction studies show that most real crystals are imperfect. Studies of the mechanical behaviour of crystalline solids have led to the recognition that some of the imperfections present are dislocations.* A principal crystal face which contains the end of a screw dislocation will appear essentially like fig. 4: i.e. it will have a step, one of whose terminations is not at the boundary, but at the end of the dislocation.

* For a review of the theory of dislocations, cf. Cottrell 1950, where dislocations of edge and screw type are defined.

In this case growth of the crystal does not eliminate the step, the process of growth consisting of a rotation of the step about the point where it joins the dislocation. The same is true for any dislocation whose Burgers vector has a component normal to the face in which the dislocation ends. Unlike the ideal crystal of n layers, this dislocated crystal consists of one layer only, in the form of a helicoid. When the dislocation is of multiple strength, we have a crystal of several interleaved helicoidal layers. When it contains several dislocations, it consists of a number (possibly one) of similar interleaved 'expanded Riemann surfaces'.

In general, the emergence of screw dislocations in a crystal face eliminates the need for surface nucleation on that face, provided only that the distance between pairs of dislocations, or between a dislocation and the boundary, exceeds the diameter of the critical nucleus of surface nucleation theory of the degree of supersaturation concerned. Suitably dislocated crystals are thus able to grow at almost arbitrarily low values of the supersaturation. The limiting case is a crystal whose faces are only twice as wide as the critical nucleus, which can just grow if it contains one screw dislocation. This is the size of the critical nucleus in the theory of three-dimensional nucleation (Becker and Döring 1935).

The theory of growth kinetics for dislocated crystals has been developed by Burton, Cabrera and Frank (1949, 1951). The main principles of the theory are outlined below.

The step in a crystal face containing the end of only one screw dislocation will run to the boundary, nearly in a straight line, when the environment is just saturated. If supersaturation is now brought about, the step will advance. It can be shown fairly easily that the rate v_0 of advance of a 'straight' step is given by

$$v_0 = 2(\alpha - 1)x_s Z \beta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where Z is the frequency with which molecules from the equilibrium vapour strike a lattice site in the surface, and β is a factor which is unity in simple cases. Qualitatively, the argument here is that all molecules which hit the surface in the 'diffusion zone' of width $2x_s$, will reach the advancing step, and there are plenty of kinks (exchange sites) to adsorb them. The number of molecules added to the step per unit length per unit time is thus $2\alpha Z x_s / a$, while the number evaporating is $2Z x_s / a$. The rate of advance is thus as given. The factor 2 arises because molecules can approach the step from above or below. The formula is derived accurately by solving for the appropriate boundary condition a differential equation of the form

$$\nabla^2 \psi - k\psi = 0$$

wherein the first term arises from surface diffusion and the second from exchange with the vapour.

The factor β may be less than unity if kinks are not close together compared with x_s , or if molecules do not always adhere to the kinks which they reach (e.g. if it is also necessary for them to arrive in the proper orientation).

Now, a curved step, whose shape corresponds to an arc of the boundary of the critical nucleus for the local supersaturation, will have a rate of advance zero. In the approximation that the critical nucleus is circular, of radius ρ_c , given by eqn. (7), the rate of advance of a step with radius of curvature ρ may be shown to be

$$v=v_0(1-\rho_c/\rho). \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Hence, the step, initially straight, advances parallel with itself, with constant velocity v_0 , over the greater part of its length, but in the neighbourhood of the dislocation end, where such advance would produce a sharp corner, it advances more slowly in such a way that its curvature is everywhere less than ρ_c^{-1} . This causes it to wind up into a spiral, centred on the dislocation end, with a spacing of approximately $20\rho_c$ between turns. The final stationary form of this rotating spiral, satisfying the differential eqn. (11) is given to a close approximation by

$$r/\rho_c + \ln(1+3^{-1/2}r/\rho_c) = 2(1+3^{-1/2})[\theta - v_0 t/2\rho_c(1+3^{-1/2})] > 0. \quad . \quad (12)$$

An approximation sufficient for most purposes is

$$r/\rho_c = 2(\theta - v_0 t/2\rho_c) > 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

The most important quantity to know is the number of turns of the spiral passing a fixed point in unit time, $v_0/4\pi\rho_c(1+3^{-1/2})$, since this quantity, multiplied by the step height, is the rate of growth of the crystal surface. Since v_0 is proportional and ρ_c inversely proportional, to the supersaturation $\alpha-1$ (when this amounts to a few per cent or less), the steady state growth rate is proportional to the square of the supersaturation.

This result is derived on the assumption that turns of the spiral are so far apart that they do not compete with each other for molecules from the vapour. This is no longer true, if they are closer together than twice the surface diffusion distance x_s ; a further increase in step density beyond this scarcely contributes to the increase in growth rate, so that the latter now becomes proportional to the first power of the supersaturation. We can now equate the rate of growth to $\beta Z d(\alpha-1)$, the same as it would be for a liquid surface. The full expression for growth rate, assuming a sufficient kink density in the steps is found to be

$$R = \beta d Z (\alpha-1) (x_s/10\rho_c) \tanh(10\rho_c/x_s), \quad . \quad . \quad . \quad (14)$$

$$= \beta d Z (\sigma^2/\sigma_1) \tanh(\sigma_1/\sigma), \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

where

$$\sigma = (\alpha-1) \sim \ln \alpha,$$

$$\sigma_1 = (10\rho_c/x_s) \ln \alpha = 10\gamma s_0/x_s kT,$$

and d is the step height. This gives a transition from the parabolic law $R = \beta d Z \sigma^2/\sigma_1$, at low values, to the linear law $R = \beta d Z \sigma$ at high values of the supersaturation.

Crude theoretical estimates of γ and x_s for iodine at 0°C , by the methods of § 2 give $\gamma a \sim 4kT$ and $x_s \sim 400a$, whence $\sigma_1 \sim 0.1$. The resulting curve

is in reasonably satisfactory agreement with the observations of Volmer and Schultze (fig. 2), when a factor $\beta=0.4$ is included. Their non-observance of departures from the linear law with phosphorus and naphthalene suggest that for these substances $x_s > 10^4 a$, which is not improbable.

§5. THE SHAPES OF GROWTH SPIRALS

The growth spiral has 'circular symmetry' (a paradoxical phrase which we shall re-examine presently) unless the factor β of eqn. (10) varies with orientation in the crystal face. An obvious source of variation is that when the step is parallel to certain close-packed directions, it may be relatively free from kinks, the kink energy w being then high. However, eqns. (4) and (5) and the accompanying estimates of energy parameters show that in general, when a crystal is growing from its vapour, the surface diffusion distance x_s will largely exceed the distance between kinks even in the step of closest-packed orientation. In this case the rate of advance of a step should not be appreciably dependent on orientation. (The molecule which misses a kink on its first approach to the step will still have several chances to find one before evaporating.)

When a crystal grows from solution, x_s probably loses its significance, and is replaced by δ , the thickness of the effectively unstirred boundary layer of solution at the crystal surface. At low temperature the distance between kinks in a close-packed step may exceed this, the rate of advance of steps in this orientation will be reduced, and polygonal spirals will result.

The segments of the polygon will only be straight at the limit of very low temperatures; in general, they will be curved convexly, but the corners will be macroscopically sharp, having essentially the same curvature as the boundary of a critical nucleus.

The 'stationary' shape of a polygonal growth spiral is defined by a variant of the Wulff construction. The ordinary version of this construction is: find the inner envelope of normals to the radius vectors of a polar diagram. It gives both the shape of the body of minimum surface free energy, from a polar diagram of surface free energy (as in eqn. (7) above), and the stationary growth shape from a polar diagram of surface growth rates. The appropriate variant of the construction here is defined as follows:

All 'corresponding' corners of the polygonal spiral are in line with one side of a polygon, which we may call the conjugate polygon: and the normal distances of a pair of straight segments of the spiral meeting at such a corner, from any point in the corresponding side of the polygon are in the ratio of their velocities of advance.

The rotational symmetry of a growth spiral may now be defined as the rotational symmetry of its conjugate polygon. It must possess at least the rotational symmetry of the crystal face, assuming a symmetrical environment.

The spacing between successive turns of a growth spiral is equal to the diameter of the critical nucleus multiplied by a numerical factor. It is, therefore, essentially inversely proportional to the supersaturation. The numerical factor is $2\pi(1+3^{-1/2})$ for spirals of circular symmetry according to eqn. (12), or 2π according to the approximate eqn. (13). It takes other values of the same order of magnitude for various polygonal spirals, e.g. four for a square spiral. 2π will serve as a representative approximate figure for all cases.

A circular or polygonal spiral step on the crystal face will appear, macroscopically, as a low cone or pyramid of vicinal faces. The slope of its faces is inversely proportional to the spacing between turns and, therefore, proportional to the supersaturation. In order of magnitude, it is about 1 minute of arc per 1% of supersaturation. We expect in general growth cones when a crystal grows from its vapour and growth pyramids when it grows from solution, with less curvature in the pyramid faces the lower the temperature.

§6. INTERACTIONS OF GROWTH SPIRALS

The theory presented above assumes only one dislocation, or two of opposite sign, to be present. The presence of more makes remarkably little difference. The detailed theory of interactions between growth spirals is discussed by Burton, Cabrera and Frank (1951). Two dislocations separated by a distance large compared with $2\pi\rho_c$ can be the centres of two independent growth spirals so that the crystal face is divided into two regions, whose growth steps proceed from the two centres. The growth rate is everywhere the same as if only one dislocation were present, but examination of the vicinal faces will reveal two growth hills. However, if the supersaturation at the centre of one is slightly greater than that at the centre of the other the territory of the second will continually shrink and, if the conditions remain constant, ultimately only one growth hill will be visible. We say that one dislocation is dominating the other. If two dislocations of the same sense are closer together than $2\pi\rho_c$ they generate a pair of non-intersecting growth spirals, thus behaving like a dislocation of double strength. They are now said to be co-operating. A co-operating pair sends out growth steps twice as fast as a single dislocation. The growth rate in the regime of low supersaturation, which gives a parabolic law, is thus doubled: this chiefly means that the regime of the linear law is reached sooner, since it is difficult to make accurate observation in the regime of the parabolic law anyway.

The chief inferences are: (a) that observations of crystal growth rate alone tell us little more about the dislocation density than whether or not there is one screw dislocation emergent in each crystal face; (b) that examination of vicinal pyramids on crystal faces does not necessarily show us all the screw dislocations emergent in the face. If we can eliminate the possibility that they are centred on specks of dirt which have settled on the face during growth, the number of pyramid apices gives a lower

limit to the number of screw dislocations emergent. Only when techniques are available that make mono-molecular steps in the crystal surface visible are we in a position to count all the screw dislocations emergent. As a rule, edge dislocations (meaning, in this context, those whose Burgers vectors lie in the plane of the crystal face) will still be invisible.

§7. OBSERVATIONAL EXAMPLES

Growth pyramids of vicinal faces have long been recognized, and were so named by Miers (1903-4). His studies and others, mainly qualitative, show general conformance to the theoretical expectations outlined above.

Growth pyramids are well seen on the major rhombohedral faces of almost any quartz crystal. The number in a face varies from one to thousands, and their size from 15 cm to fractions of a millimetre. There is a specimen in the Field Museum in Chicago, in which a single vicinal pyramid, with very flat faces, occupies the whole crystal face, about 15 cm across; in some natural crystals, and almost all artificial quartz crystals, the growth pyramids are numerous, and have distinctly curved faces.

Much more information can be obtained when we can see the edges of crystal monolayers. The first to do so knowingly was Griffin (1950), who studied surfaces of natural beryl crystals, by ordinary microscopy (later using phase contrast microscopy with advantage). In this case the step height is only 8 Å, and the visibility of the step is probably due to localized natural etching of some kind (fig. 5, Plate I).

A much simpler case is provided by the long chain paraffin crystals ($C_{36}H_{74}$) grown from solution in light petroleum, studied first by Dawson and Vand (1951) (figs. 6 and 7, Plates II and III). As the step height for a monolayer of these crystals is about 50 Å, it is easily visible by direct electron microscopy after metal shadowing. The examples shown (obtained in Bristol University by H. F. Kay and Mrs. B. J. Appelbe) are selected for their simplicity showing one and two emergent dislocations. Other crystals with larger numbers of dislocations are also seen.

Silicon carbide provides another convenient case for observation. Spiral markings which may be seen with a low power magnifier, or the naked eye, have been known for some time (Menzies and Sloat 1929). These often represent steps several hundred Angström units high. Verma (1951) and also Amelinckx (1951) have recently found that much smaller steps can be made visible by phase contrast microscopy on lightly silvered SiC surfaces, and that these correspond to the unit cell thickness, which is 15 Å for the common variety. The examples shown (figs. 8-11, Plates IV and V) illustrate various points in the theory described above—a detailed discussion of them has been given by Frank (1951). A point of particular interest is that the dislocations in SiC are often visibly hollow. This can be generally expected with a dislocation of Burgers vector exceeding about 10 Å (Frank 1951 a) unless it can easily dissociate into partial dislocations. Another is that the remarkable polytypism of carborundum

(which has variant structures with repeating lattice units of up to 198 layers) is immediately comprehensible when we learn that the crystals frequently grow upon dislocation groups of large total strength. The repetition period of the crystal structure is determined simply by the total dislocation strength of the dislocation, or dislocation group, which dominates the growth. A peculiarity in the crystal stacking may thus recur with perfect regularity at intervals far exceeding the conceivable range of direct influence of interatomic forces. Examples of this were hitherto extremely puzzling (Ramsdell 1947, Honjo 1950).

Crystal growth based on dislocation groups of multiple strength, producing large growth steps (of height 80 Å upwards) but otherwise essentially similar to any other kind of crystal growth based on dislocations, can be followed while in progress by techniques developed by Forty (1951). He has made studies of CdI_2 and PbI_2 growing from aqueous solution, photographing successive stages of the rotation of the growth spirals (figs. 12–15, Plate VI). These spirals were first seen, with poorer resolution, in Bunn and Emmett's cinematograph film of crystal growth (cf. Bunn and Emmett 1949). In favourable cases the step height can be measured with a precision of about 1 Å, by use of interference fringes formed between the front and back surfaces of a thin crystal, while still immersed in solution. Fig. 16 (Plate VII) is an example of a photograph taken with this technique, and serves to show that the central dislocation passes right through the crystal making steps of equal height on each face. One face only is growing and has its step wound into a spiral. The other face is close to the glass and so screened from the supersaturated solution: its step remains straight and radial. The equality of step height on both faces can be seen from the equality of intensity in the interference pattern in two quadrants at each crossing. Crystals may also be examined dry. Fig. 17 (Plate VIII) is an example. It is remarkable that virtually all peculiarities noticeable in photographs of growth patterns on SiC can be matched with similar examples on CdI_2 .

A notable observation is that dislocations penetrating the basal plane are commonly absent initially in PbI_2 crystals, and are formed in a catastrophic event, after considerable lateral growth has occurred. Until then the crystal grows as an extending plate with no increase of thickness whatever. The most probable explanation is that non-uniform distribution of impurities stresses the thin plate to its true yield limit, when it gives way by buckling and shearing to produce dislocation groups usually of large total Burgers vector. It is possible that growth at the edges is by the classical Volmer mechanism, and does not involve dislocations.

§8. THE ORIGIN OF DISLOCATIONS IN CRYSTAL GROWTH

The origin of the initial dislocations in a crystal presents no great problem. Either the crystal begins to grow on a foreign solid, which may be dislocated already, or its nucleation occurs at a very high supersaturation, which need only be a little higher to nucleate a dislocated

crystal instead of a perfect one. Or, if dislocations are not created immediately, the probable mode of early growth is dendritic, and when the dendritic arms unite they are likely to do so in imperfect registration with each other, so producing dislocations. If growth then continues at low supersaturation, there is a natural selection of suitably dislocated crystals.

This, however, accounts only for one or a few dislocations emergent near the centre of each growing face. Elsewhere, such crystals would be substantially perfect. While such crystals apparently exist, observations suggest that in the majority of crystals dislocations occur more or less throughout their volume. The writer is inclined to attribute this principally to deformations caused by impurities. We shall not give here a full discussion of such phenomena, but we may sketch an approach to the subject.

(a) If a solution contains two components capable of forming a mixed crystal, and $G_L(x)$, $G_s(x)$ are the free energies of liquid and solid as a function of molar fraction x , then if the tangent to $G_L(x)$ at the actual liquid composition x_L cuts the curve $G_s(x)$ at two points, deposition can occur on crystal seeds within a finite range of compositions. The decrease of free energy is greatest for a composition x_s^* defined by

$$G_s'(x_s^*) = G_L'(x_L), \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where primes signify differentiation with respect to x . If there is only one seed of composition x_0 , and the temperature is adjusted so that deposition just occurs, the composition of the deposit x_D is given by

$$\frac{x_D - x_0}{x_s^* - x_0} = \frac{\alpha}{\alpha + \beta}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

where

$$\alpha = G_s'', \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

(higher derivatives being neglected) and

$$\beta = 2\alpha'VE/(1-\nu). \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

α' is the derivative of lattice parameter with respect to x , V is molar volume, E Young's modulus and ν Poisson's ratio. β is typically 10^{12} ergs. An order of magnitude for α is obtained by supposing it to arise only from entropy of mixing, whence:

$$\alpha = RT/x(1-x), \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

so that α is 10^{13} ergs when $x=1/4000$. α also contains temperature-independent contributions. With decrease of temperature it can approach zero and become negative. The mixed crystal is then unstable, and will segregate if solid diffusion can occur; but the composition of the deposit from solution is stably defined so long as $(\alpha + \beta)$ remains positive.

(b) If the substrate crystal is not homogeneous (and it cannot remain so during growth according to (a)), the appropriate value of x_0 is some

mean value defined by the lattice parameter at the surface. Its determination involves the solution of an elastic problem. In general, it is not the same at all points of the crystal surface.

(c) With appropriate variation of composition, it is possible to have curved crystal lattices, free from dislocations, and as free from elastic stress as is possible in any kind of mixed crystal (i.e. when the local stress fields around individual atoms are averaged out). For example, the curved lattice generated by geometrical inversion of the simple cubic lattice, in which all lattice planes become spheres passing through the origin, still has almost perfectly cubic unit cells (with angular errors equal to lattice parameter divided by local radius of curvature). The lattice parameter is proportional to the square of the distance from the origin, so that this curved lattice can be occupied in a mixed crystal system allowing 5% variation of lattice parameter throughout a spherical shell whose thickness is $2\frac{1}{2}\%$ of its radius, or 5% of the radius of curvature of the most highly curved lattice surfaces. A crystal bounded by lattice surfaces would show, qualitatively, the kind of curvature commonly observed in crystals of dolomite, $(\text{Ca}, \text{Mg})\text{CO}_3$ (cf. Heddle 1923). Non-cubic systems allow a greater variety of curved lattice than we have described here.

(d) When α is small compared with β , the crystal which deposits from solution will approximate to one of the unstressed curved forms mentioned under (c) and will be not much less stable in one of these than in the rectilinear lattice. It will grow on in the extrapolation of whatever curved lattice it has to begin with. However, it is characteristic of curved lattices that, since there is a limit to the variation of lattice parameter, the region in which they can be occupied is bounded. To grow outside this region the crystal must degenerate into a 'lineage structure' of roughly parallel overgrowths upon the curved core, or alternatively, undergo plastic deformation of the core. In the first case the junctions of the overgrowths are the sites of dislocations.

(e) Plastic deformation of the core may proceed by displacement of dislocations already present. Unpublished observations by Griffin suggest that this may have occurred in the growth of Beryl crystals. Crystals which grow initially as very thin plates (CdI_2 , PbI_2 , hypothetically $\text{Si}(\text{'})$) make a special case in that this deformation can occur in a catastrophic manner by buckling, so producing the very large dislocation groups and high growth steps which are observed.

(f) Much has been omitted from this account. An important point is that the composition of the liquid continually changes during deposition. What matters here is the composition of the liquid in contact with the crystal, which will become constant if there is no stirring. This agrees with the observation that crystals grown upwards from a melt are in general much more free from lineage structure than those grown downwards. Solid recrystallization eliminates stirring in the highest degree and is known to produce crystals which preserve lattice parallelism over great distances (Guinier 1949).

§9. THE RANGE OF APPLICATION OF THE PRESENT THEORY

We may say briefly that the dislocation theory of crystal growth is likely to be appropriate to all cases in which the crystal is observed to grow as a convex polyhedron, especially at low supersaturation. We have every reason to believe the correctness of surface nucleation theory for growth without dislocations, at relatively high supersaturation. It is likely that in the case of complex or dissociable molecules, and of substances growing from solution, that the critical degree of supersaturation needed for this process to operate at an appreciable rate will show a wider range of variation than is indicated in the elementary derivation of the theory for growth from a vapour of simple molecules. Instead of being the order of magnitude 50% it may perhaps sometimes come down to a few per cent. Nevertheless it is believed that the same theory in principle applies generally, for growth from solution as well as from vapour. Now, whenever diffusion in the medium is of importance (i.e. in solution, or in the presence of a foreign gas) it must be expected that the surface nucleation will occur at the corners of the crystal (where diffusion will maintain the highest degree of supersaturation, if growth is proceeding). Then, except with a delicate adjustment of the supersaturation, dendritic branching from the corners is very liable to occur. Thin plates, however, will be more stable in this respect than equant bodies.

At supersaturations too low for surface nucleation it is possible that foreign particles, of an appropriate substance, settling on the crystal surface, can help to initiate new layers. This also should fail below some finite critical supersaturation. The author does not know of observations which enable one to assess the extent and importance of this phenomenon.

Cases may be clearly recognized in which growth is promoted by the persistent re-entrant angles in the surface provided by twinned crystals. Good examples are provided by fluorite and calcite (Frank 1949 a). This too should presumably fail below a finite critical supersaturation which may, however, be very low. There should be no danger of failing to recognize when this mode of growth promotion is in operation.

We arrive thus at the conclusion stated in the first sentence of this section.

Now we must consider the important case of growth from a melt. As we remarked in § 2, Burton and Cabrera find that various crystal faces 'melt' at various temperatures—i.e. cease to be atomically plane save for small numbers of defects—but that for crystals in contact with their vapours a family of faces, sufficient to bound the solid, will usually remain 'unmelted' up to the actual melting point. Contact with a liquid will lower the energy of surface defects, and therefore lower the temperature of surface melting. It follows that when a solid is in contact with its own liquid, and in equilibrium, it may happen that none of its surfaces is rightly described as atomically plane in the first approximation. In consequence, theory in its present state does not allow us to assert that a perfect crystal cannot grow from its melt at an arbitrarily low

degree of supercooling. What we now need to know is the time which the solid-liquid surface requires to reach its equilibrium configuration—a problem of kinetics in a co-operative system which it will be difficult to treat either accurately or with convincing approximations. Perhaps, even though the equilibrium surface is atomically rough when in equilibrium, the process of growth fills in the depressions, and produces a flat surface, so that surface nucleation is still required unless dislocations are present. The degree of supercooling needed for surface nucleation would be expected to be small in this case. This idea fits the observed fact that crystal growth of face-centred cubic metals from slightly supercooled melts is in the form of dendrites, branching in the directions of the octahedral corners. When the melt is not supercooled, but heat is removed through the solid, the freezing surface shows little crystallographic dependence, but is virtually a smooth isothermal surface: the undoubted presence of dislocations suffices to account for this, whether or not there exist surfaces which would be slow-growing in the perfect crystal.

The growth of an etch pit is equivalent to the growth of a crystal, and should be promoted by the presence of a dislocation in the same way: with the interesting difference that whereas the 'line tension' of a dislocation draws it away from a salient corner, so that one dislocation will promote growth on one face only, it holds it at a re-entrant corner, so that a single dislocation should be able to promote the development of three or four faces in an etch pit.

There are various processes of solid-solid reactions which constitute crystal growth—oxidation for example. Whenever the bounding surface is a crystallographic plane either of the growing or the shrinking crystal, we may infer that the presence of dislocations is important, and that the theories of this article are essentially applicable. Otherwise, e.g. in typical grain boundary migrations in metals, they are not.

REFERENCES

- AMELINCKX, S., 1951, *Nature*, **167**, 939.
 BAUMHAUER, H., 1915, *Zeits. Krist.*, **55**, 249.
 BECKER, R., and DÖRING, W., 1935, *Ann. Physik*, **24**, 719.
 BUNN, C. W., and EMMETT, H., 1949, *Disc. Far. Soc.*, **5**, 119.
 BURTON, W. K., and CABRERA, N., 1949, *Disc. Far. Soc.*, **5**, 33, 40.
 BURTON, W. K., CABRERA, N., and FRANK, F. C., 1949, *Nature*, **163**, 398 ; 1951, *Phil. Trans. Roy. Soc. A* **243**, 299.
 COTTRELL, A. H., 1949, *Progress in Metal Physics*, I. Edited by B. Chalmers (London: Butterworths Scientific Publications), Ch. 2.
 DAWSON, I. M., and VAND, V., 1951, *Nature*, **167**, 476 ; 1951, *Proc. Roy. Soc. A* **206**, 555.
 FRANK, F. C., 1949, *Disc. Far. Soc.*, **5**, 48, 67 ; 1949 a, *Ibid.*, **5**, 186 ; 1951, *Phil. Mag.*, [7], **42**, 1014 ; 1951 a, *Acta Cryst.* (in the press).
 FORTY, A. J., 1951, *Phil. Mag.*, [7], **42**, 670.
 FRENKEL, J., 1945, *J. Phys. U.S.S.R.*, **9**, 392 ; 1946, *Kinetic Theory of Liquids* (Oxford: Clarendon Press).

- GIBBS, J. W., 1878, *Collected Works*, 1928 (New York : Longmans, Green and Co.) Vol. I, p. 325, footnote.
- GRIFFIN, L. J., 1950, *Phil. Mag.*, [7], **41**, 196.
- GUINIER, A. J., and TENNEVIN, J., 1949, *Philips Res. Rep.*, **4**, 316.
- HAWARD, R. N., 1939, *Trans. Far. Soc.*, **35**, 1401.
- HEDDLE, M. F., 1923, *The Mineralogy of Scotland*, (St. Andrews : University Press). Vol. I, Plate XLVII, fig. 271.
- HONJO, G., MIYAKE, S., and TOMITA, T., 1950, *Acta Cryst.*, **3**, 396.
- KOSSEL, W., 1927, *Nachr. Ges. Wiss., Göttingen*, **1927**, 135.
- MENZIES, A. W. C., and SLOAT, C. A., 1929, *Nature*, **123**, 348.
- MIERS, Sir H. A., 1903, *Proc. Roy. Soc.*, **71**, 439 ; 1904, *Phil. Trans. Roy. Soc. A* **202**, 459.
- RAMSDELL, J. S., 1947, *Amer. Mineralogist*, **32**, 64.
- STRANSKI, I. N., 1928, *Z. Phys. Chem.*, **136**, 259.
- VERMA, A. R., 1951, *Nature*, **167**, 939 ; *Phil. Mag.*, [7], **42**, 1005.
- VOLMER, M., 1939, *Kinetik der Phasenbildung* (Dresden and Leipzig : Steinkopff).
- VOLMER, M., and SCHULTZE, W., 1931, *Z. Phys. Chem. A* **156**, 1.





THE MATHEMATICAL WORKS OF JOHN WALLIS, D.D., F.R.S.

by

J. F. SCOTT, Ph.D., B.A.

"His work will be indispensable to those interested in the early history of The Royal Society. I commend to all students of the Seventeenth Century, whether scientific or humane, this learned and lucid book."—Extract from foreword by Prof. E. N. da C. Andrade, D.Sc., Ph.D., F.R.S.

Recommended for publication by University of London

12/6 net

Printed and Published by

TAYLOR & FRANCIS, LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4.



"IF ITS WORTH
KEEPING
BIND IT"

Pat. No. 463268

EASIBIND

The Self-Binding Device illustrated is ideal for both PERMANENT and TEMPORARY binding of technical journals. They allow the pages to open quite flat, and parts are easily inserted by means of steel wires (supplied with the binders) and can be removed and replaced at any time.

Arrangements have been made to supply these Binders (to take up to 4 parts of the Advances in Physics') to Libraries, Museums and Subscribers at the inclusive cost of

(Gold blocked with Title)

15/-

(Quarter-bound Rexine)

Applications to be made to the Publishers :—

TAYLOR & FRANCIS, Ltd., Red Lion Court, Fleet St., London, E.C.4.

Annals of Science

A QUARTERLY REVIEW OF
THE HISTORY OF SCIENCE
SINCE THE RENAISSANCE

EDITORS

D. McKIE, D.Sc., Ph.D.,
University College, London.

HARCOURT BROWN,
M.A., Ph.D.,
Brown University, Providence, R.I.,
U.S.A.

H. W. ROBINSON,
Former Librarian,
Royal Society of London.

ANNUAL SUBSCRIPTION

£2 0s. 0d.

OR

10s. 6d.

PER PART
POST FREE



Contents of Vol. 7, No. 3, September 1951

PETER SHAW AND THE REVIVAL OF CHEMISTRY. By F. W. GIBBS,
M.Sc., Ph.D. (PLATES XXIV AND XXV.)

**THE HISTORY OF PHYSIOLOGY AT ST. BARTHOLOMEW'S HOSPITAL,
LONDON.** By JOHN L. THORNTON, A.L.A.

**ENGLISH TREATMENT OF THE RELATIONSHIP BETWEEN THE RISE
OF SCIENCE AND THE RENAISSANCE, 1740-1840.** By HERBERT
WEISINGER, A.M., Ph.D.

**ASTRONOMICAL AND CHRONOLOGICAL CALCULATIONS AT
NEWMINSTER IN 1428.** By LYNN THORNDIKE, A.M., Ph.D.

CHEMISTRY IN ROZIER'S JOURNAL.—IV AND V. By E. W. J. NEAVE,
M.C., M.Sc., Ph.D. (PLATES XXVI AND XXVII.)

REVIEWS